2006

Security in heterogeneous wireless networks

Vijay Bulusu
Louisiana State University and Agricultural and Mechanical College

SECURITY IN HETEROGENEOUS WIRELESS NETWORKS

A Thesis

Submitted to the Graduate Faculty of
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Master of Science in Systems Science

in

The Department of Computer Science

by
Vijay Bulusu
B.E. in Information Technology, Indian Institute of Information
Technology - Calcutta, 2004
August 2006

Acknowledgments

I would like to thank my advisor Dr. Arjan Durresi for his support, encouragement and patience during my graduate study. This work would not have been possible without his insightful comments and suggestions. It was a stroke of good fortune that lead to my joining his research group and an interesting research area. I would be eternally grateful to him for his endless supply of interesting research problems; for helping me think about research from a broader perspective and all his down-to-earth advice.

I offer my sincere gratitude to my committee members Dr. Jianhua Chen and Dr. Bijaya Karki for their cooperation.

I would like to thank all my colleagues and friends, especially Vamsi Paruchuri (soon to be Dr. Vamsi Paruchuri) for all his thoughtful suggestions and support.

Last, but not least, I would like to gratefully acknowledge the Department of Computer Science, Louisiana State University for providing me the necessary resources for completing this thesis.

Table of Contents

Abstract

The proliferation of a range of wireless devices, from the cheap low power resource starved sensor nodes to the ubiquitous cell phones and PDA's has resulted in their use in many applications. Due to their inherent broadcast nature Security and Privacy in wireless networks is harder than the wired networks. Along with the traditional security requirements like confidentiality, integrity and non-repudiation new requirements like privacy and anonymity are important in wireless networks. These factors combined with the fact that nodes in a wireless network may have different resource availabilities and trust levels makes security in wireless networks extremely challenging.

The functional lifetime of sensor networks in general is longer than the operational lifetime of a single node, due to limited battery power. Therefore to keep the network working multiple deployments of sensor nodes are needed. In this thesis, we analyze the vulnerability of the existing key predistribution schemes arising out of the repeated use of fixed key information through multiple deployments. We also develop SCON, an approach for key management that provides a significant improvement in security using multiple key pools. SCON performs better in a heterogeneous environment. We present a key distribution scheme that allows mobile sensor nodes to connect with stationary nodes of several networks.

We develop a key distribution scheme for a semi ad-hoc network of cell phones. This scheme ensures that cell phones are able to communicate securely with each other when the phones are unable to connect to the base station. It is different from the traditional ad hoc networks because the phones were part of a centralized network before the base station ceased to work. This allows efficient distribution of

key material making the existing schemes for ad hoc networks ineffective. In this thesis we present a mechanism for implementing authenticated broadcasts which ensure non-repudiation using identity based cryptography. We also develop a reputation based mechanism for the distributed detection and revocation of malicious cell phones. Schemes which use the cell phone for secure spatial authentication have also been presented.

Chapter 1
Introduction

Security is wireless networks is harder than the wired networks because of their inherent broadcast nature. This is due to the fact that every packet transmitted in a wireless network can be intercepted by all nodes within the communication range of the transmitter. This underlying vulnerability not only makes security in wireless networks extremely important but also adds additional requirements like privacy, anonymity and resource optimization.

1.1 Classification of Wireless Networks

There are many different types of wireless networks like sensor networks, ad hoc networks, vehicular networks and cellular networks. Our work mainly involves All these networks have different charactersistics and constraints because of which protocols specific to particular networks have to be developed. Many of the networks that we envisage for the future are heterogeneous where nodes of different characteristics and constraints would combine to serve many different applications. Some of the characteristics which differentiate between nodes are battery power, communication range, computational power and memory. We discuss the characteristics of the wireless nodes and networks for which we have designed the protocols.

1.1.1 Wireless Sensor Networks

A Distributed Sensor Network consists of a large number of autonomous, self-organizing sensors with limited battery power, computational power, communication range and memory. These nodes communicate through the wireless medium. Each node is equipped with integrated sensors, data processing capabilities and

short-range radio communications. Sensor Networks can be used in a variety of applications like military sensing and tracking, environmental monitoring, patient monitoring and tracking, smart environments, Disaster Management etc. The sensor nodes are deployed in large numbers in or close to the phenomenon [5]. Traditionally, these nodes sense the physical environment and communicate the data to a base station. The base stations process the data and take appropriate action.

1.1.2 Wireless Sensor Actor Networks

Many applications like protection from forest fires, chemical attacks, military surveillance, home automation [6, 46] etc are sensitive to the time delay between the sensing of a phenomenon and performing a necessary action. These have lead to the emergence of a new class of heterogeneous networks called Wireless Sensor Actor Networks (WSAN). Here the sensors relay the information about the location and the intensity of the phenomenon to the actors and the actors perform the necessary actions.

1.1.3 Semi Ad hoc Cellular Networks

A cellular phone, also known as cell phone, mobile phone is a telecommunication device with the capability of a conventional telephone. These are portable wireless devices and connect to the network through RF communication. Due to their low cost and multitude of features, these phones have been transformed from expensive equipment used for business to a low cost personal item. It is estimated that there are over 2 billion cell phones worldwide [1]. These phones typically have low power transceivers which typically transmit data and voice up to a few miles where the mobile tower (base station) is located. This base station connects the cellular phone to the backbone telephone network. The mobile phones can not communicate when they are unable to connect to the base station[10].

In this thesis, we propose to use cellular phones as ad hoc networks when they are unable to connect to the base station. We assume that the phones function normally when they are able to connect to the base station but they function as the nodes of an ad hoc network when they are unable to connect to the base station. We also assume that the phones can connect to both the base station and the ad-hoc network at the same time. Ad hoc networks consist of self configurable mobile nodes which communicate through the wireless medium. These networks do not require a centralized base station and nodes may communicate through multi-hop routing. The nodes are free to move randomly which makes the topology highly dynamic and unpredictable[60]. The ability to change from centralized mode to ad hoc mode allows these networks to leverage the advanatges of both systems. No requirement for fixed infrastructure and the easily available mobile phones allow rapid deployment. This makes them ideal for emergency situations like natural or human-induced disasters. People can use the omni-present mobile phones to connect to these networks. We refer to networks which convert from a centralized system to ad hoc in the middle of their deployment as semi ad hoc networks.

1.2 Motivation

Many of the applications for which these networks have been envisaged require security. Key distribution is the most critical part of security establishment because it has to be achieved in an unsecured environment.

1.2.1 Security in Sensor Networks

There are many requirements which any key distribution scheme must satisfy. Some of these requirements are

- **Non-Deterministic Deployment:**The network administrator has no control over the placement of sensor nodes in the deployment field. The nodes are

scattered in the deployment region and the probability of any two nodes being neighbors after deployment is equal. These nodes discover their neighbors and setup communication paths and start sensing, processing and communicating without any manual intervention. This is the preferred approach for deployment because it is scalable for the large number of nodes in the network. This approach is used when a large number of nodes are being deployed or the deployment is being done in inaccessible terrain. For e.g. military applications like surveillance in hostile areas and disaster management.

- **Self Configuring:** This is an extension of the previous requirement. Once the nodes are deployed, they should be able to communicate with each other and establish the distributed security protocol without any manual intervention. The protocols should be able to account for the fact that all nodes may not be deployed at the same time.

- **Heterogeneous Nodes:** A sensor network consists of a large number of nodes. A network is said to be homogeneous when all the nodes of the network are of the same type. On the other hand a network with many different nodes is called a heterogeneous network. Coordination between the various nodes is much easier in a homogeneous network. Moreover, all nodes of the network can perform all the jobs, which makes node deployment much simpler. With the proliferation in the types of wireless devices, heterogeneous networks is a fact of life. Ideally, all the protocols should be developed in a way that they benefit from the heterogeneity of nodes. For e.g. tasks which require more power can be assigned to the few nodes with greater battery power.

Our work addresses two problems related to the security in wireless networks.

- **Secure Continuity in Sensor Networks :** There could be a vast difference in the time duration for which a phenomenon needs to be monitored and the lifetime of sensor nodes. This makes the secure addition of new nodes to an existing network essential for the network to survive. The secure addition of nodes is complicated because different nodes of the network expend resources at different rates. Nodes closer to the base station are required to route a lot more messages and may die quickly. When new nodes are added to the network there may be a significant number of previously deployed nodes which are alive. If the previously deployed nodes are discarded, the problem of secure network continuity becomes trivial because every new deployment could use a different keypool. Secure addition of new nodes to an existing network would also be useful in situations where a part of the network is destroyed because of some natural or man-made causes.

  One way to reduce the operation costs of sensor networks is to reuse the nodes of previous deployments should not be discarded because after an initial overhead of establishing secure connectivity, the number of nodes in the network would increase. This would provide better connectivity and the increase the lifetime of the nodes because the load of routing is distributed over a greater number of nodes. Even if the previously deployed nodes are discarded, they continue communicating with each other. As a result they would contend for the common medium along with the newly deployed nodes resulting in more collisions. The retransmission of lost packets would lead to the wastage of battery power. Moreover, if there is a time gap between two deployments then these active nodes of previous deployments could continue sensing and transfer any data to the base station once the new nodes are deployed.

Many of the existing key predistribution schemes [13, 19, 22, 34, 61] can be used for multiple deployments, but the security offered by these schemes decreases with time. They suffer because of the repeated use of fixed key information. The capture of each node increases the fraction of keys known by the adversary. When a certain number of these nodes are captured, the adversary has enough keys to compromise a large number of links making the network ineffective. Addition of new nodes to the network with keys from the same key pool will not help because the keys in the new nodes are already compromised. We define this as the Secure Continuity problem.

- **Secure Mobility Across Multiple Networks:** All existing schemes make use of the same key pool for stationary and mobile nodes. Although this approach works fine when the mobile nodes are restricted to one network, they fail when the mobile nodes need to move through multiple networks a great geographical distances. The use of the same key pool in all networks is not possible because the capture of nodes in one scheme would compromise the secure links established in other networks.

1.2.2 Security in Semi Ad hoc Networks of Cell Phones

The capabilities of cell phones have increased dramatically over the last few years. In addition to the standard telephone features, the phones also Instant Messaging, MMS, Internet access etc. More advanced features like music and video streaming, digital camera, document scanner are being bundled with the cell phone [7]. These features have transformed the cell phone from a simple phone to a digital swiss army knife.

- **Non-Repudiation in Semi-Ad hoc Networks:** There are many applications in which cellular phones can be used by for real-time communication

between large numbers of people. It could be used by the law enforcement agencies to transmit urgent messages to people in subway stations and football games. People traveling on high speed highways can receive messages about accidents on their phones ahead of time to manage traffic and relive congestion. Using cellular phones in these situations can be used to communicate with a large number of people when no other means of communication are available. In natural disasters like hurricanes where no other means of communication are available, cellular phones can be used to communicate important messages. People trapped inside buildings could use their mobile phones to alert others in case of fires or floods. These applications require the data to be transmitted to all the users hence data confidentiality is not a requirement. In this case secure broadcast authentication is required to achieve guaranteed non-repudiation.

There are another set of applications for mobile phones in semi ad-hoc mode where data needs to be sent from one user to the other even when the phones are unable to connect to the base station. This would greatly help in emergency situations where the base stations may not work. Allowing the mobile phones to communicate when the base station is dysfunctional could help in coordinating the relief effort in situations like hurricanes. Even in situations where the base stations are working, there are locations where connectivity to the cellular network is low. In those situations the ability of the phones to communicate in the ad hoc mode can be useful.

A phone which is out of the reach of a base station may be allowed to connect to it through another phone. If some base stations near the ad hoc network of cell phones are functional, they can be used to send instructions to all

the nodes. Information flow may start from the base station and through the phones connected to the base station end up in the phones of the ad hoc network.

- **Trust in Semi Ad-hoc networks:** The ease with which a user can communicate with a large number of other users makes this system vulnerable to attacks ranging from pranks to terrorist attacks. Terrorists could artificially increase the population of an area by asking all other cell phone user to go there before a terrorist attack. To ensure that the perpetrator of these attacks is identified, these schemes require secure broadcast authentication to achieve guaranteed non-repudiation. This approach is reactive because it identifies the attacker after the attack. Clearly this is not enough for some attacks that may are possible with this system. Protocols need to be developed which would prevent a malicious cell phone user from misguiding other users[9].

- **Secure Spatial Authentication using Cell phones:**

  There are many applications in wireless networks where access is granted to a user only when the user is located in certain predefined locations [26, 40, 55]. For eg. a doctor should be able to access the medical records only when he is located inside the hospital and not in cafeteria. In this scenario the doctor has access to the medical records only when he is located in a safe place like his office and not in a public place like the cafeteria. There are other scenarios where the cell phone can be used as an access card to enter into buildings. Using the cell phone for this purpose helps in minimizing the number of such cards which need to be carried by a user and also allows easy access revocation.

1.3 Contribution

The main contributions of this thesis are

- Our work identifies the weaking of security over successive deployments for the existing key predistribution schemes for sensor networks [13, 19, 22, 34, 61]. We also develop scalable resource efficient protocol called Secure Continuity in Sensor Networks (SCON) to address this secure continuity problem.

- We develop two schemes which allow a mobile sensor node to move across multiple disjoint networks of statioanry sensor nodes.

- Our work proposes to use cell phones in the ad hoc mode for emergency communication which can be used in disaster management. We propose to use the cell phones in the semi ad hoc mode (refer to section 1.1.3). We also propose a key distribution scheme that ensures non-repudiation in this operating environment.

- We propose a distributed reputation management scheme for cell phones in the ad hoc mode. Our scheme proposes a novel distributed approach for node revocation in an ad hoc network of cell phones.

- We develop a scheme which can provide Secure Spatial Authentication using dual cell phones (refer to section 1.2.2).

1.4 Structure of the Thesis

The rest of the thesis is organized into 5 main chapters. Chapter 2 presents our protocol to address the problem of secure continuity in sensor networks. This protocol works in both homogeneous and heterogeneous networking environments. Chapter 3 presents our scheme that allows mobile sensor nodes to connect with

the stationary nodes of multiple networks. Chapter 4 presents a scheme to establish secure broadcasts and point-to-point communication in a semi ad hoc network of cell phones. In Chapter 5 we present a reputation based scheme for distributed node revocation in a semi ad hoc network of cell phones. Chapter 6 we present a protocol which uses dual cellular phones for secure spatial authentication. We conclude in Chapter 7.

Chapter 2

Secure Continuity in Wireless Sensor Networks

2.1 Introduction

A Distributed Sensor Network consists of a large number of autonomous, self-organizing sensors with limited battery power, computational power, communication range and memory. These nodes communicate through the wireless medium. Each node is equipped with integrated sensors, data processing capabilities and short-range radio communications. Sensor Networks can be used in a variety of applications like military sensing and tracking, environmental monitoring, patient monitoring and tracking, smart environments, Disaster Management etc. The sensor nodes are deployed in large numbers in or close to the phenomenon [5]. Traditionally, these nodes sense the physical environment and communicate the data to a base station. The base stations process the data and take appropriate action. Many applications like protection from forest fires, chemical attacks, military surveillance, home automation [6, 46] etc are sensitive to the time delay between the sensing of a phenomenon and performing a necessary action. These have lead to the emergence of a new class of heterogeneous networks called Wireless Sensor Actor Networks (WSAN). Here the sensors relay the information about the location and the intensity of the phenomenon to the actors and the actors perform the necessary actions. The compromise of these networks may result in serious consequences. Security in Sensor Networks is non-trivial because of limited resource and the possibility of physical node capture. Carman, Kruus and Matt have analyzed the security constraints for sensor networks [11]. There are many types of routing attacks to which the sensor networks are vulnerable[30].

This chapter deals with secure key distribution in sensor networks which is one of the toughest aspects of security because it has to be accomplished in an unsecured environment. Traditional schemes like Kerberos [32, 42], which rely on trusted third party infrastructure, are infeasible for sensor networks with a large number of nodes and limited communication range. The limited resources make the use of Asymmetric Cryptosystems infeasible because they require lots of computation and memory.This makes algorithms like Diffie-Hellman key agreement [18] and RSA [48] undesirable. Moreover, symmetric key ciphers and hash functions are two to four orders of magnitude faster than digital signatures [11]. Asymmetric cryptosystems cannot be used even to establish session keys because that would leave the nodes vulnerable to Denial of Service attacks [11, 12, 57].

A popular technique for key distribution for nodes with limited resources; little or no deployment knowledge is key predistribution [12, 13, 19, 22, 34, 35, 61]. This involves the loading of information required by the nodes for key establishment before their deployment. But, when a node is physically captured all the keys present in that node are known to the adversary. This not only compromises the links established by the captured nodes but also compromises links between uncompromised nodes.

The nature of sensor networks is such that it is almost impossible to know which nodes would be within communication range of each other after deployment. There are applications where some nodes are more likely to be neighbors than others. In such cases the keys in each node can be decided apriori [19, 34]. Even if the exact position of nodes after deployment is known, the large number of nodes makes precise deployment infeasible. Our approach works for cases with and without deployment knowledge.

There could be a vast difference in the time duration for which a phenomenon needs to be monitored and the lifetime of sensor nodes. This makes the secure addition of new nodes to an existing network essential for the network to survive. The secure addition of nodes is complicated because different nodes of the network expend resources at different rates. Nodes closer to the base station are required to route a lot more messages and may die quickly. When new nodes are added to the network there may be a significant number of previously deployed nodes which are alive. If the previously deployed nodes are discarded, the problem of secure network continuity becomes trivial because every new deployment could use a different keypool. Secure addition of new nodes to an existing network would also be useful in situations where a part of the network is destroyed because of some natural or man-made causes.

One way to reduce the operation costs of sensor networks is to reuse the nodes of previous deployments should not be discarded because after an initial overhead of establishing secure connectivity, the number of nodes in the network would increase. This would provide better connectivity and the increase the lifetime of the nodes because the load of routing is distributed over a greater number of nodes. Even if the previously deployed nodes are discarded, they continue communicating with each other. As a result they would contend for the common medium along with the newly deployed nodes resulting in more collisions. The retransmission of lost packets would lead to the wastage of battery power. Moreover, if there is a time gap between two deployments then these active nodes of previous deployments could continue sensing and transfer any data to the base station once the new nodes are deployed.Many of the existing key predistribution schemes [13, 19, 22, 34, 61] can be used for multiple deployments, but the security offered by these schemes decreases with time. They suffer because of the repeated use of fixed key information. The

capture of each node increases the fraction of keys known by the adversary. When a certain number of these nodes are captured, the adversary has enough keys to compromise a large number of links making the network ineffective. Addition of new nodes to the network with keys from the same key pool will not help because the keys in the new nodes are already compromised.

In this paper we present SCON, a security management approach to handle key distribution by using a separate key pool for each phase of node deployment. The nodes of different deployments securely communicate through the use of special nodes called bridge nodes. These nodes have keys from the key pool of the current deployment and the previous deployment. The number of keys from each key pool is based on the fraction of nodes from the current deployment and the previous deployments in the deployment region. To the best of our knowledge SCON is the first attempt to address the problem of security through multiple deployments. SCON is independent of any scheme and can be implemented along with several existing schemes [13, 19, 22, 34, 35, 61]. The use of separate key pools for each deployment results in reduced connectivity over multiple deployments because nodes of different deployments share no keys. Connectivity is improved by using path-key establishment where any two nodes which are within the communication range of each other can establish a secure direct link if there exists a secure path between them. Although SCON works with just the normal sensor nodes, we propose the use of nodes with different capabilities for better performance in terms of connectivity and overhead.

SCON works best when nodes of different capabilities are available. We have used three types of nodes for our scheme, which are normal nodes, strong nodes and actors. The normal nodes are the resource-starved nodes discussed earlier. The strong nodes have more computational power, memory and battery power when

compared to the normal nodes as a result of which they hold more keys. Similar nodes have been used for routing purposes in [53]. SCON proposes to use these strong nodes as bridge nodes. The actors on the other hand are robots that have the ability to move. They have significantly more memory, computational power and battery power than normal nodes. They are primarily deployed to perform actions based on the data sensed by the nodes of the network [6].We propose to use the actors to improve the security and connectivity in the network. Our scheme would work even without the actors although their use improves the performance of the scheme. These actors can communicate among themselves and the base station using asymmetric cryptography. Our scheme offers better connectivity at lesser overhead with all these different nodes.

We make the assumption that the lifetime of a phase of node deployment is significantly lesser than lifetime of the network. The compromise of nodes is assumed to be random and the attacker captures a fixed fraction of total deployed nodes in each phase of deployment. When a node is compromised, it is assumed that the attacker is able to read all the information in the node(including the keys). The attacker also assumed to have the ability to change the data and programs in the node. In this paper we assume that the capture of all nodes including bridge nodes is equally likely. We would like to emphasize that the capture of strong nodes is possible although our scheme ensures that the adversaries would not be able to specifically target them. Our simulations also assume that same number of nodes are deployed in each phase of deployment although the scheme can be extended for other cases. In our approach we assume a secure base station but all the other nodes of the network are vulnerable to capture. Since the actors need to communicate with nodes of different deployments it gets the keys from the base station when it needs them. This is done to minimize the compromise of keys when an

actor is captured. To detect the capture of an actor,schemes for node audition like [50] may be used.

We compare the performance of Random Key predistribution scheme [22], q-composite random key predistribution scheme [13] and Random key predistribution with deployment knowledge [19] with SCON and without SCON. Our simulations show that the security of these schemes after using SCON is twice as good after three deployments.

The remainder of the paper is organized as follows. The existing schemes over which our approach has been implemented are summarized in Section 2.2. We present our approach in Section 2.3 and the analysis and simulations in Section 2.4.

2.2 Background Work

To the best of our knowledge SCON this is the first attempt to address the problem of managing network continuity through multiple deployments having as goal the tradeoff among security, connectivity and network cost. SCON can be implemented along with most of the existing predistribution based schemes like [13, 19, 22, 34, 35, 61]. Eschenauer and Gligor proposed the random key predistribution scheme [22], also known as the basic scheme. It is based in the interesting properties observed in random graphs. A random graph $G(n, p)$ is a graph of n vertices with p being the probability of any two vertices having an edge. For monotone properties there exists a value of p such that the probability of the graph being connected moves from "non-existent" to "certainly true" [52]. It consists of key predistribution, shared key discovery and path key establishment. The key predistribution phase involves the generation of a key pool of S keys. Each node selects m keys randomly from the key pool S and stores them in its memory. After deployment, shared key

discovery is performed. Here all pairs of neighboring nodes try to find a common key between them. If such a key exists then it is used to setup a session key between the nodes. This is followed by path key establishment where secure links are established between neighboring nodes that do not share keys but have a secure path between them. This improves the connectivity of the network and reduces the overhead of communication at the cost of a one-time overhead of key establishment.

The basic scheme is further improved by Chan, Perrig and Song [13]. It is similar to the previous scheme except that any two nodes need to have q common keys ($q > 1$) to establish a secure link. As a result the keys required by the attacker to compromise a link increases. This scheme would establish fewer secure links when compared to the basic scheme because the probability of two nodes sharing more than one key is less than the probability of them sharing one key. This problem is addressed by the reduction of the key pool size. With a reduced key pool size the capture of a node compromises a larger fraction of the key pool making it less secure. The interplay of these opposing factors result in the scheme working better than the basic scheme at lower levels of compromise but the performance of this scheme dips as the level of compromise increases.

The most important information that can benefit key predistribution is the knowledge of nodes that are likely to be neighbors after deployment. No such information is assumed in the above schemes. The scheme presented by Du, Deng, Han, Chen and Varshney [19] uses this knowledge to improve security and connectivity. In this scheme the total number of sensors to be deployed is divided into multiple groups with each group having a deployment point. All these deployment points are arranged in a grid. The probability of two nodes from different groups being neighbors decreases with the increase in distance between their deployment points. Since the basic scheme assumes no deployment knowledge, every node should pick

keys from the same key pool because all the nodes are equally likely to be neighbors. With deployment knowledge the probability of any two nodes being neighbors is not the same. Nodes from the same group and nearby groups are more likely to be neighbors than the others. Therefore, when two groups are far away from each other, their key pools could be different. This scheme is inherently more secure than the basic scheme because a larger number of keys can be used to provide the same level of connectivity. As a result, the capture of a node leads to the compromise of a smaller fraction of the key pool. The connectivity provided by this scheme is also superior to the basic scheme because the probability of any two neighboring nodes sharing keys is higher.

There are schemes presented by Liu and Ning [34] which use bivariate polynomials for key establishment. In this scheme each node is given shares of different polynomials through which it can establish a secure link with another node which has a share from the same polynomial. The polynomial can be reconstructed with t shares. These schemes become vulnerable over time with the capture of nodes. All the above schemes suffer because they reuse the information for key distribution. The capture of each node gives the attacker a greater part of the fixed key information. When the number of nodes captured crosses a threshold value, the number of secure communication links becomes low and the network becomes virtually dead. Addition of new nodes to the network will not help because the attacker knows the keys in these nodes.

Our approach SCON solves the problem of weakening security over time and multiple phases of deployments by using different keys for different deployments. This would keep the fraction of the links compromised in each deployment constant, thereby preventing the total number of secure links from becoming low.

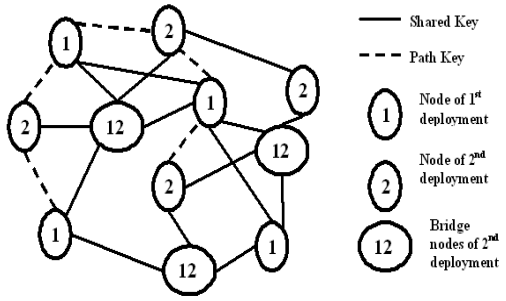


FIGURE 2.1. This figure shows the working of SCON for two deployments. Nodes of 1^{st} and 2^{nd} deployments establish shared keys with the bridge nodes. Once this is done two neighboring nodes of different deployments which share a link with a common bridge node establish a path-key

2.3 SCON: Secure Management of Continuity

We solve the problem of continuity through the use of a separate key pool for each phase of deployment. The interconnectivity of these nodes is made possible through the use of special nodes called bridge nodes. The bridge nodes have keys from the previous phase of deployment along with the keys from the current deployment. As a result they are able to establish secure links with the newly deployed nodes as well as the previously deployed nodes. The degree of interconnectivity between the nodes of various deployments is dependent on the number of bridge nodes. The greater the number of bridge nodes, the greater is the probability of two nodes from different deployments establishing a secure link.

We implement our approach on the random key predistribution scheme [22] for clear illustration. Like all other predistribution-based schemes our scheme has three stages, which are key information predistribution, shared key discovery and path-key establishment [22].

- **Key information predistribution:** This stage is performed offline before the nodes are deployed. Separate disjoint key pools S_1,S_2.....S_n and key identifiers for deployments 1,2.....n are generated. Each node of the i^{th} deployment is given m keys from key pool S_i. The bridge nodes of the i^{th}

deployment have keys from the pool S_i along with keys from S_{i-1}. The number of keys in a bridge node from each of these key pools depends on the likeliness of the node from that deployment being a neighbor of a node of the i^{th} deployment.

- **Shared Key discovery:** In this stage each node discovers the nodes within its communication range with which it shares keys. This can be done by making each node broadcasts the list of key identifiers of the keys in its memory as soon as it is deployed. All the neighbors of these nodes compare the list of the key identifiers broadcasted with their own key identifiers. Any two nodes share a key if they have some common key identifier(s). Using this approach the nodes of the i^{th} deployment establish a secure network among themselves because they have keys from the same key pool S_i. The bridge nodes of the i^{th} deployment are able to connect to the nodes of the i^{th} deployment and the nodes of the $(i-1)^{th}$ deployment. In this stage the strong nodes could be detected because they would broadcast the key identifiers from the key pools of two different deployments. This vulnerability can be fixed by making the normal nodes broadcast some random key identifiers so that both the strong nodes and the normal nodes broadcast equal number of key identifiers.

- **Path-key establishment:** In this phase a session key is established between those nodes which are within communication range of each other and do not share a common key but have a secure path between them. Once a path-key is established, the nodes can communicate directly without any intermediate nodes.

The bridge nodes deployed in the i^{th} deployment get connected to the nodes of the i^{th} and $(i-1)^{th}$ deployment through shared key discovery because the bridge nodes have keys from S_i and S_{i-1}. The bridge node may have other neighbors that belong to deployments prior to the $(i-1)^{th}$ deployment. The nodes of the $(i-1)^{th}$ deployment would be connected to such nodes. This allows the bridge nodes to connect to all nodes of previous deployments through path-keys. This keeps the network connected across nodes of multiple deployments.This is clearly illustrated by Figure 2.1.

The use of strong nodes in SCON would improve connectivity. The presence of actors improves the connectivity even further. The performance of our approach improves with the use of these nodes although the approach would work even without them.

2.3.1 Strong Nodes as Bridge Nodes

Bridge nodes are responsible for the interconnection of the nodes of different deployments. This can be achieved when the bridge node is able to establish enough secure links with the nodes of the current deployment and those of the previous deployments. The ability to hold more keys will certainly be helpful for the bridge nodes. This makes a strong case for the use of strong nodes as bridge nodes. This would allow the bridge nodes to have more keys that will in turn allow them to establish more links with the nodes from both the current and the previous deployment. The use of strong nodes as bridge nodes leads to SCON performing better.

During the shared key discovery phase when all nodes broadcast their key identifiers, the attacker can differentiate between the normal nodes and the strong nodes because the strong nodes would broadcast more key identifiers. With this

the attackers could capture all the strong nodes and prevent nodes of multiple deployments from being connected. This vulnerability can be fixed by providing the normal nodes with randomly selected key identifiers from the previous deployment so that the total number of key identifiers in the normal nodes and the strong nodes is equal.

2.3.2 Actor Nodes

Actors are resource rich nodes with a larger memory; better processing capabilities and infinite battery power along with the ability to move. There could be one or more actors in the network and as always more the better [6]. While we make use of actors we do not count on their presence. We look at some ways the actors can help in improving the connectivity and security of the network while performing their other duties.

Key Establishment

The number of the bridge nodes is usually small when compared to the total nodes deployed. There might be regions in the deployment area where the nodes of the current deployment may not get connected to the nodes of the previous deployments. This situation can be improved through the use of actors. These actors can go to the regions with low connectivity and establish path keys between nodes that are left unconnected even after path-key establishment.

The use of actors would add another stage to SCON called the Actor key establishment. This stage has to be performed after key information predistribution, shared key discovery and path-key establishment.

- **Actor Key Establishment:** Each node in the network shares a unique key with the base station. In this stage actor(s) goes into regions in the deployment area where the connectivity between the nodes of various deployments

is not good and establishes links between nodes. If the actor wants to establish a session key between two nodes that are not connected after the path-key establishment, it sends their ids to the base station. The base station responds by sending the actor the unique keys that it shares with the nodes. The actor can communicate with the two nodes using these keys and establish a secure link between them. This approach of storing no sensor key information in the actor protects the network if an actor is captured.

Actor key establishment also causes lesser overhead on the normal nodes of the network because the actor, which has infinite battery power, does a part of the key establishment.

Node Audition

The actors can also audit the nodes of the sensor network by using the method SWATT proposed by Seshadri, Perrig, Doorn and Khosla [50]. It works on the assumptions that the base station can generate the memory map of a sensor node based on the nodes in its neighborhood and that the hardware of the nodes is not modified by the adversary. If the memory map of the node is different from the copy generated by the base station, it is assumed that the adversary has compromised the node and he has loaded his own malicious program into the node. The base station has a verification procedure, a copy of which is loaded in the memory of each node. This method is used to determine whether a node is compromised or not. The keys in the compromised nodes are not used in the subsequently deployed bridge nodes. As a result the subsequently deployed bridge nodes have fewer compromised keys thereby improving connectivity.

- **Actor based node audition:**The actor creates a random challenge and sends it to the node being audited. The node computes the response to the

challenge using the verification procedure and the random challenge sent by the actor and returns it to the actor. The actor then compares it with the possible values for that challenge obtained from the base station. If the two values do not match the node is considered compromised.

The base station can also audit the actor using this method to ensure that the actor is not compromised. The presence of actors results in the scheme performing better.

2.4 Analysis and Simulations

We define interconnection as the ability of a bridge node to establish a secure connection with two nodes of different deployments. For the simulations and the derivations we assume that all bridge nodes are strong nodes, they hold twice as many keys as the normal nodes and if the bridge node is deployed with the i^{th} deployment, it will have half its keys from the i^{th} deployment and remaining keys will be from the $(i-1)^{th}$ deployment. The two main metrics for our approach are security and connectivity. We also analyze the overhead for our approach.

2.4.1 Analysis of SCON

Security

Intuitively our approach is more secure than the existing schemes [13, 19, 22, 34, 35, 61] over multiple deployments because the number of keys deployed increases with each phase of deployment. The fraction of keys compromised because of the capture of a node decreases as a result and security is improved.

Let m_i be the number of keys in each node of the i^{th} deployment. Let S_i be the key pool for the i^{th} deployment with $|S_i|$ being the number of keys in pool S_i. c_{ij} is the number of nodes captured in the i^{th} phase of deployment with j being the phase in which the captured node was deployed. The capture of a bridge node is equivalent

to the capture of two nodes i.e. one each of the i^{th} and the $(i-1)^{th}$ deployments. Let P is the probability of a key being compromised. In the first deployment c_{11} nodes are compromised. The probability of a key not being captured is: $\left(1-\frac{m_1}{|S_1|}\right)^{c_{11}}$. During the second deployment c_{21} nodes from the first deployment and c_{22} nodes from the second deployment. The probability of a key not being compromised after the second deployment is: $\left(1-\frac{m_1}{|S_1|}\right)^{c_{11}}\left(1-\frac{m_2}{|S_1US2|}\right)^{c_{21}}\left(1-\frac{m_2}{|S_1US_2|}\right)^{c_{22}}$. Hence, after i deployments the probability of a key being compromised is:

$$P = 1 - \prod_{i=1}^{n}\prod_{j=1}^{i}\left(1-\frac{m_i}{\left|\bigcup_{k=1}^{i} S_k\right|}\right)^{c_{ij}} \tag{2.1}$$

Equation (2.1) clearly shows that the probability of a key being compromised is least when the key pools are disjoint. If we use the same key pool again and again, then $|S_1 \cup S_2 \cup \ldots \cup S_n| = |S_1|$. In this case the expression would reduce to

$$P = 1 - \left(1-\frac{m_1}{|S_1|}\right)^{x} \tag{2.2}$$

where x is the total number of nodes captured at a given point in time. In case of a fixed key pool, the capture of nodes over multiple deployments is analogous to the capture of all the nodes at once.

Situations can be envisaged where the adversary is interested in compromising a particular link of the network. This is analogous to compromising of a certain key. Equation (2.2) can also be used to calculate the number of nodes required to be captured to compromise a given link with a certain probability. This can be done by fixing the value of P and finding n. If the key required by the adversary is from the key pool S_i, then n is the number of nodes which have the keys from S_i. For a given rate of node capture calculating the time in which n nodes from a given

deployment would be captured is trivial.

Connectivity

The nodes of the i^{th} deployment connect among themselves because they all share keys from the key pool S_i. These nodes are also able to connect to the bridge nodes because they also contain keys from S_i. The bridge nodes also contain keys from the pool S_{i-1} that lets them connect to the nodes of the $(i-1)^{th}$ deployment. They allow the bridge nodes of the i^{th} deployment connect to the other nodes of the previous deployments through path-keys. Having the necessary bridge nodes is critical to achieve the desired level of connectivity.

Let c be the fraction of the bridge nodes out of the n nodes of the i^{th} deployment. Let A be the area of the deployment region. The density of the nodes is $\rho = N/A$. Let r be the communication range of a node. Assuming that the range of the sensor node is uniformly circular, the area of communication for a node within one hop is πr^2 . The number of nodes within communication range from a node of the i^{th} deployment is $\rho\pi r^2$. Let us assume that the degree of bridge nodes required for a node be d_b. The fraction of bridge nodes for the given degree of bridge nodes is given by:

$$c = \frac{d_b}{\rho\pi r^2} \tag{2.3}$$

Equation (2.3) gives us the number of nodes required when the required degree of bridge nodes is fixed. We now derive an expression for the probability of there exists at least one secure path from a node of the i^{th} deployment to a node of the $(i-1)^{th}$ deployment.

Let P be the probability of the node A of i^{th} deployment and the node B of the $(i-1)^{th}$ deployment having a secure path through a bridge node. Let M be a bridge

node and P_a, P_b be the probability of having secure paths between nodes A, M and B, M respectively. The probability of there being a secure path between A and B through M is $P_a.P_b$. Therefore the probability that there is no secure path from A to B through M is $(1 - P_a.P_b)$. If there are d bridge nodes in the neighborhood of both A and B, the probability of there being no secure path between A and B would be $(1 - P_a.P_b)^d$. Hence the probability that there is at least one secure path between A and B is:

$$P = 1 - (1 - P_a.P_b)^d \tag{2.4}$$

The probabilities P_a and P_b are fixed and they depend on the schemes over which our approach is implemented. By fixing the value of P, we can obtain the value of d, which gives us the total number of bridge nodes.

Let $|S|$ be the size of the key pools and m be the number of keys per node of the i^{th} and the $(i-1)^{th}$ deployments. As per our assumption the strong nodes have $2m$ keys (m each from S_i and S_{i-1}). For the q-composite random key predistribution:

$$P_a = \frac{\begin{pmatrix} |S| \\ q \end{pmatrix} . \begin{pmatrix} |S| \\ 2\,(m-q) \end{pmatrix} . \begin{pmatrix} 2\,(m-q) \\ (m-q) \end{pmatrix}}{\begin{pmatrix} |S| \\ m \end{pmatrix}^2} \tag{2.5}$$

Let x nodes of the $(i-1)^{th}$ deployment be captured. The capture of one node means that the probability of a key being compromised is $m/\,|S|$. The probability of a key not being compromised is $1 - m/\,|S|$.The probability of a key not being compromised after the capture of x nodes is $(1 - m/\,|S|)^x$. Let $|S'|$ be the number of keys from the key pool of the $(i-1)^{th}$ deployment which are not compromised.

$$|S'| = |S| \cdot \left(1 - \frac{m}{|S|}\right)^{x} \tag{2.6}$$

If a constant rate of node capture is assumed, P_b will be less than P_a because by the time nodes of the i^{th} deployment are deployed, some nodes of the $(i-1)^t h$ deployment would have been captured. The expression for P_b is:

$$P_b = \frac{\begin{pmatrix} |S'| \\ q \end{pmatrix} \cdot \begin{pmatrix} |S| \\ 2(m-q) \end{pmatrix} \cdot \begin{pmatrix} 2(m-q) \\ (m-q) \end{pmatrix}}{\begin{pmatrix} |S| \\ m \end{pmatrix}^2} \tag{2.7}$$

The expressions for P_a and P_b are derived for q-composite scheme. By substituting $q = 1$ we obtain the expressions for the basic random key predistribution scheme.

Overhead for Key Establishment

The random predistribution scheme[22] shows that a node is able to establish a secure link with almost all of its neighbors within 3 hops of path-key establishment. Similarly, a bridge node deployed in the i^{th} deployment connects to all its neighbors in the $(i-1)^{th}$deployment within three hops. The nodes of the $(i-1)^{th}$ deployment are already connected to nodes of all previous deployments. As a result the overhead for a bridge nodes of the i^{th} deployment to connect to all nodes before the $(i-2)^{nd}$ deployment is the same. For a node in the i^{th} deployment to get connected to all the nodes of the previous deployments, it has to get connected to the bridge nodes in its neighborhood. Through the bridge nodes the nodes of they can get connected to the nodes of the previous deployments.

Let A be a node of the i^{th} deployment and B be a node of the $(i-1)^{th}$ deployment. Let M be a bridge node with keys from S_i and S_{i-1}. $L(i,j)$ be the path-length

for key establishment from a node of the i^{th} deployment to a node of the j^{th} deployment. The total overhead of key establishment for all deployments between i and j would be:

$$L(i,j) = L(i,i-1) + L(i-1,j) + (i-j).c \tag{2.8}$$

Here $i > j$ and c is a constant.The constant is included in the equation because the after the nodes of the i^{th} deployment get connected to the nodes of the $(i-1)^{th}$ deployment, connecting to any other node after that requires a constant overhead.

Key Exclusivity

If two nodes share one or more keys exclusively, then their communication is invulnerable to the capture of any number of nodes[29]. It is the probability of a communication link between any two neighboring nodes would be invulnerable to the capture of any number of other sensor nodes. In the case of random predistribution, the probability of two nodes possessing a particular key is $(m/|S|)^2$ where m is the number of keys in each node and $|S|$ being the total number of keys in the key pool. If n is the total number of nodes then the probability of a key is invulnerable i.e. it exists only in two neighoring nodes is $\left(\frac{m}{|S|}\right)^2 \left(1-\frac{m}{|S|}\right)^{n-2}$. After k deployments the total number of nodes deployed would increase to kn. Let IV^{rand} and IV^{SCON} be the link invulnerability for random key predistribution with and without SCON. The probability of the link invulnerability would be

$$IV^{rand} = \left(\frac{m}{|S|}\right)^2 \left(1-\frac{m}{|S|}\right)^{kn-2} \tag{2.9}$$

For SCON, let the fraction of bridge nodes per deployment be c. Let S_i be the key pool of the i^{th} deployment. Keys from the pool S_i would be present in the nodes of the i^{th} deployment and the bridge nodes of the $(i+1)^{th}$ deployment. Nodes of other deployments baring from the bridge nodes of the $(i+1)^{th}$ would not have keys

from S_i. As a result the link invulnerability would be independent of the number of deployments. The probability of link invulnerability after k deployments would be

$$IV^{SCON} = \left(\frac{m}{|S|}\right)^2 \left(1 - \frac{m}{|S|}\right)^{n+c-2} \tag{2.10}$$

Equations 2.9 and 2.10 clearly show that the probability of link invulnerability in Random Key predistribution unlike SCON gets worse with each deployment. This is due to the fact that the number of nodes with keys from the same key pool increases in random key predistribution. With SCON, this probability remains constant because the keys from any keypool S_i are reused only for the nodes of the i^{th} deployment and the strong nodes of the $(i+1)^{th}$ deployment.

2.4.2 Simulations

In this section we compare Random key predistribution scheme, q-composite random predistribution scheme and random key predistribution with deployment knowledge with SCON and without SCON. Our simulations show that these schemes achieve much better security with SCON. We also simulate the performance of the scheme with and without strong nodes. While the security achieved by SCON with or without strong nodes is similar, the connectivity is much better with the use of strong nodes.

Our simulation considers a square deployment area of 200x200m^2 with the communication range of each sensor and the strong node being 20m of distance. We assume all links to be symmetric meaning that if node A is within the communication range of node B then node B is in the communication range of node A. The capture of nodes by the adversary leads to the compromise of keys. We assume the size of the key pool to be 10000 with each node having 75 keys. The capture of 1

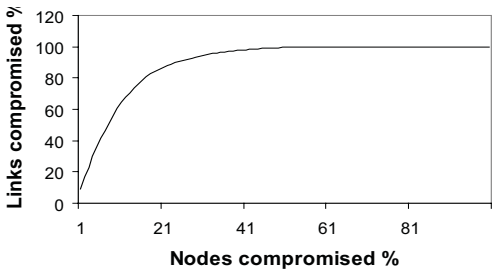
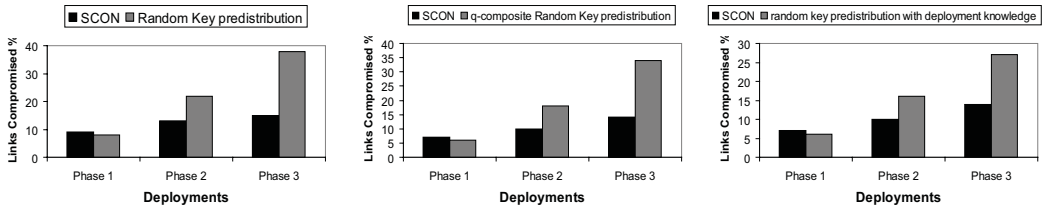


FIGURE 2.2. Fraction of links revealed per nodes compromised in random predistribution

node (0.1% of the nodes) results in the compromise of 75 keys (0.75% of the keys). Initially the capture of the nodes reveals a greater portion of the key pool. The subsequent capture of nodes reveals keys, which are already known to the adversary. Our simulations show that the capture of 20% nodes reveals the compromise of 85% links. This trend is clearly shown in Figure 2.2.

We have implemented SCON on the random key predistribution scheme, q-composite random predistribution and random key predistribution with deployment knowledge. We compare the fraction of the links compromised when a fixed number of nodes are captured (1% of the total nodes deployed). The results in Figure 2.3 show a much lower number of links compromised for the capture of every node when the above schemes are implemented with SCON. The use of multiple disjoint key pools ensures that the capture of a node reveals a lesser fraction of the total keys deployed resulting in the compromise of fewer nodes.

The overhead for key establishment decreases with the use of nodes of varying capabilities because the interconnection between various deployments is better. We calculate the average number of hops for path-key establishment for a node to establish a secure link with all its neighbors. Figure 2.4 clearly shows that the number of hops for path-key establishment decreases with the availability of better nodes. The actors can establish a path-key between any two nodes within



(a) random key predistribution (b) q-composite random key predistribution (c) random key predistribution with deployment knowledge

FIGURE 2.3. Fraction of links Compromised in random key predistribution, q-composite random predistribution and random key predistribution with deployment knowledge with and without SCON for multiple deployments with a random node capture of 1% of the total nodes deployed. The fraction of bridge nodes is 20% of the nodes deployed.

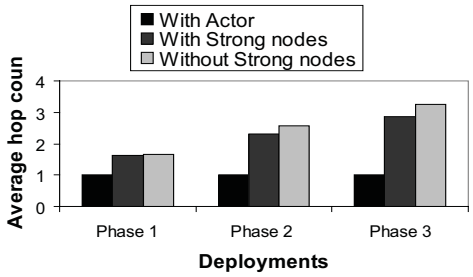


FIGURE 2.4. Comparison of the one time overhead for key establishment using SCON on random key predistribution with actors, strong nodes and normal nodes

communication range in one hop. The actors can be used in cases where this overhead for key establishment becomes very large. We also compare the one time overhead for key establishment in basic random key predistribution with or without SCON in Fig.5. In this case the overhead without SCON is lesser because although a significantly larger fraction of links are compromised, the basic random scheme establishes a larger number of links in the first place. In Figure 2.4 and Figure 2.5 the assumption is made that each node is aware of all the keys that have been compromised in the previous deployment. This assumption has not been made previously.

Deploying an appropriate number of bridge nodes is critical to the performance of SCON. In Figure 2.6 we use strong nodes as bridge nodes. An increase in the

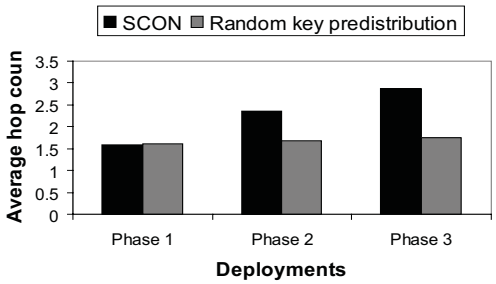
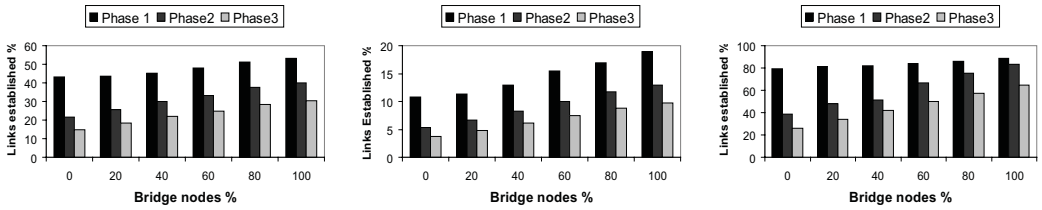


FIGURE 2.5. Comparison of the one time overhead for key establishment with and without SCON on random key predistribution

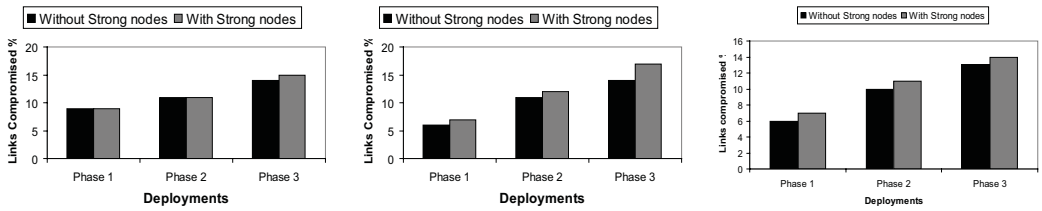


(a) random key predistribution (b) q-composite random key predistribution (c) random key predistribution with deployment knowledge

FIGURE 2.6. Fraction of links established to the nodes within communication range in random key predistribution, q-composite random predistribution and random key predistribution with deployment knowledge with SCON for varying values of the fraction of bridge nodes. Here strong nodes are used for bridge nodes

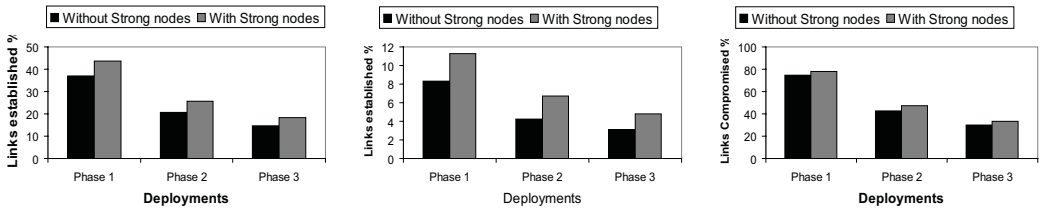
number of bridge nodes results in some nodes with more keys which end up establishing more keys in the shared key discovery phase. If the number of bridge nodes is too few then the other nodes will not be able to connect to the bridge nodes. More bridge nodes result in an increase in the number of links established in shared key discovery phase which reduces the overhead caused by path-key establishment. We have observed that the fraction of the links compromised remains nearly same with the increase in the number of bridge nodes. This is because the compromise of a bridge node reveals twice as many keys as the capture of a normal node would yield.

We compare the performance of SCON with and without strong nodes in Figure 2.7 and Figure 2.8. The security of SCON does not depend on the use of strong



(a) random key predistribution (b) q-composite random key predistribution (c) random key predistribution with deployment knowledge

FIGURE 2.7. Compares the security offered by SCON over random key predistribution, q-composite random predistribution and random key predistribution with deployment knowledge for multiple deployments with and without strong nodes.



(a) random key predistribution (b) q-composite random key predistribution (c) random key predistribution with deployment knowledge

FIGURE 2.8. Compares the connectivity offered by SCON over random key predistribution, q-composite random predistribution and random key predistribution with deployment knowledge for multiple deployments with and without strong nodes. Here connectivity is expressed as the ratio of the no. of nodes with which links are established to the total no. of nodes within communication range

nodes but depends on the use of separate key pools for each deployment. As a result the schemes provide similar security with and without strong nodes. The strong nodes have more keys because of which they are able to establish more links which improves the connectivity. Having strong nodes only improves the connectivity and not security.

Chapter 3

SUMO: Secure Key Distribution in Mobile Heterogeneous Sensor Networks

3.1 Introduction

Sensor Networks can be used in a variety of applications like military sensing and tracking, environmental monitoring, patient monitoring and tracking, smart environments, Disaster Management etc. The sensor nodes are deployed in large numbers in or close to the phenomenon [5]. These nodes typically sense the physical environment and send relevant data to a base station. In many applications like protection from forest fires, chemical attacks, military surveillance, home automation [6, 46] etc the use of mobile sensor nodes is fundamental. The nodes themselves may not move, but may be placed on the mobile objects which move in the network. For e.g. sensor nodes on a mobile tank of hazardous chemicals would communicate with other sensor nodes in case of a leak. To detect and extinguish forest fires, a sensor node may be placed on a fire truck which would interact with other stationary sensor nodes on the ground and guide the truck to the exact location of the fire. Several other military applications can be thought of where these mobile nodes could be very useful. Another possible application is navigation using these networks. We envisage many applications where people could navigate through sensor networks using common omnipresent devices like cellular phones. For e.g. a man stuck in a building on fire may use his cellular phone to interact with the stationary sensor networks deployed in the building to find the best escape route. All these problems can be modelled as mobile nodes interacting with stationary nodes in a sensor network.

All existing schemes make use of the same key pool for stationary and mobile nodes. Although this approach works fine when the mobile nodes are restricted to one network, they fail when the mobile nodes need to move through multiple networks a great geographical distances. The use of the same key pool in all networks is not possible because the capture of nodes in one scheme would compromise the secure links established in other networks. To address these problems we propose two schemes for secure key predistribution between the stationary nodes and the mobile nodes of the sensor networks. The first scheme uses a separate key pool for links between mobile and static nodes. From this key pool the mobile and stationary nodes randomly select m keys and $e(e << m)$ keys respectively. Having fewer keys in the stationary nodes ensures that the capture of a stationary node compromises a small fraction of the mobile key pool. The second scheme uses a large key pool which is segmented into smaller key pools. All the nodes of a particular stationary network select m keys randomly from one of the small segments whereas the mobile nodes select m nodes from the entire key pool. It is ensured that the probability that a mobile node would have some keys from each of the segments is high.

We analyze the performance, merits and demerits of both these schemes and compare their performace through mathematical analysis and simulations. Both schemes assume secure well connected stationary networks. To minimize overhead and increase security our scheme does not attempt to connect the mobile node with all the stationary nodes. Our scheme instead allows the mobile node to communicate with some(not all)stationary nodes from all points in a network. This is ensured by the unequal sharing of keys between the mobile and stationary nodes.

Our schemes are designed to minimize the compromise of secure links mobile and stationary nodes by the capture of both stationary and mobile nodes. The capture

of a mobile nodes in the scheme that uses a separate mobile key pool would compromise a larger portion of the key pool than the capture of a stationary node because the mobile nodes have a more keys from the mobile key pool. In the segmented key pool based scheme, the compromise of a mobile node would compromise a small portion of the keys from each segment of the key pool. Therefore, the total number of keys compromised when a mobile node is captured is lesser than the separate mobile key pool based scheme. The stationary nodes are deployed in hostile inaccessible regions whereas the mobile nodes are typically deployed of objects of importance. Based on this we believe that the capture of mobile nodes is much harder than the capture of a stationary node. Also the number of mobile nodes is going to be considerably less than the stationary nodes.

The remainder of this chapter is organized as follows. Section 3.2 discusses some existing schemes in current literature which are relevant to our scheme. We then present the two schemes in section 3.3. Section 3.4 has all the mathematical analysis, simualations and comparison of the two schemes.

3.2 Previous Work Done

To the best of our knowledge this is the first attempt at developing a key distribution scheme for mobile nodes with unrestricted mobility over multiple sensor networks. This scheme assumes that the networks of stationary nodes are well connected using any of the existing schemes. We now discuss some of the existing schemes that are relevant to the our schemes. All the work based on which our schemes have been designed has been described in section 2.2.

Our objective is to allow the mobile nodes to have the ability to operate in multiple sensor networks each of which would have keys from a separate key pool. This would make the existing schemes ineffective because the mobile nodes would

be able to operate in only a small portion of the network. Schemes that assume deployment knowledge face the same problem and hence can not be used with mobile nodes. We address this problem by using a different key pool for connecting mobile nodes to static nodes and also by using disjoint segments of a large key pool for static nodes and the whole key pool for the mobile nodes. Our schemes achieves this with very little memory overhead on the stationary nodes of the network.

3.3 Our Schemes

Mobile nodes operate in multiple networks of stationary nodes. When a mobile node moves into a particular network of stationary nodes it interacts with them. This paper presents two schemes which are able to establish secure links between the mobile nodes and stationary nodes of a sensor network. Our schemes ensures confidentiality and integrity of the messages transmitted between the mobile and stationary nodes. Both these schemes minimize the storage overhead on the stationary nodes of the network. We also present a tradeoff between security and connectivity for both our schemes. We now describe the two schemes in detail.

3.3.1 Separate Key Pool Scheme

In this scheme we use a separate key pool to connect mobile nodes with the stationary nodes of the network. Each mobile node randomly selects some keys from this key pool. All stationary nodes also select some keys from this key pool randomly before they are deployed. The number of keys from the mobile key pool in each mobile node is far greater than the number of these keys in each stationary nodes. This not only reduces the overhead on stationary nodes but also reduces the number of keys compromised when stationary nodes are captured. The advantage of this scheme is that the communication between mobile and stationary nodes is independent of the key distribution scheme used to securely connect the stationary

network. We divide our scheme into different stages which are key predistribution, key discovery and location key establishment. We now present each of these stages briefly

- **Key predistribution:** This stage is performed before the nodes are deployed. A mobile key pool S of size $|S|$ is generated along with the key identifiers. All mobile and stationary nodes are given m and e ($e << m$) keys from S respectively.

- **Session Key discovery:** When a mobile node wants to talk to the stationary nodes of the network, it broadcasts the list of its key identifiers. The static nodes match the list of broadcasted identifiers with their own identifiers. If a static node shares a key with the mobile node it establishes a secure session key with the mobile node. The mobile nodes may establish more than one links (if possible) to increase redundancy and reliability.

- **Location key establishment:** Once a mobile node establishes a session key at a particular location, it can store the key in its memory. Whenever the node visits that particular location again, it could reuse the session key. This would make session key discovery a one time overhead. This would make key Discovery for a location, a one time overhead. If the overhead of storing keys at all the locations is high, the mobile nodes could store the session keys for only the frequently visited locations.

Having fewer keys in stationary nodes reduces the probability of a mobile node sharing a key with a particular stationary node. But, we assume that the density of the stationary nodes in the deployment region is high. As a result the probability that the mobile node would establish secure links with some of its stationary

neighbors is high. The mobile nodes can communicate with all the stationary nodes of the network through these nodes.

3.3.2 Segmented Key Pool Scheme

The idea behind this scheme is to give the mobile nodes a small number of keys from the key pools of all the stationary networks with which the mobile nodes may interact. The number of keys from each of the key pools may depend on the frequency with which the mobile node visits a particular network. The number of keys from the key pool of a stationary network in a mobile node is much less than the number of nodes that are present in a stationary node. Like the previous scheme we leverage on the fact that a mobile node has several stationary nodes in its communication range at any point inside the network. Even though the probability of a mobile node sharing a key with a particular stationary node is small, the probability of sharing a common key atleast some nodes in its neighborhood is high. This allows the mobile nodes to interact with the stationary nodes without any memory overhead on the static nodes.

Like the previous scheme using Separate key pools, this scheme also has three stages which are key predistribution, session key discovery and location key discovery. In Key predistribution we generate a large key pool S if size $|S|$. This pool is divided into segments $S_1, S_2...S_n$ and one of these segments is assigned to each sensor network. All of the static nodes of a network i randomly select m keys from the key pool S_i. The mobile nodes on the other hand randomly select m keys from S. The session and location key discovery stages are exactly same as in the case of Separate key pools.

On an average the fraction of keys in a mobile node from a particular segment S_i is $1/n$. It is not mandatory that all n segments be in use. Some segments can

be kept for future deployments. This makes the segmented approach extremely flexible. We compare the two schemes extensively in the next section.

3.4 Analysis and Simulations

The metrics for the analysis of this scheme are

- **Security:** It is the probability of a secure link between a mobile and stationary being compromised with the capture of a node.

- **Connectivity:** It is the probability of a mobile node establishing q secure links with the stationary nodes from any point in the network.

- **Overhead:** It is the memory overhead on the stationary nodes to store the keys of the mobile key pool.

3.4.1 Mathematical Analysis

In this section we look at the performance of Key predistribution using separate and segmented key pools using the metrics discussed above.

Key Predistribution using Separate Key Pool

In this scheme we have a separate mobile key pool from which the mobile and stationary nodes randomly select keys. These keys are used to establish secure links between the stationary nodes and the mobile nodes. Let e be the number of keys from the pool S in each stationary node. We analyze the situation where a mobile node needs to establish q secure connections from a point in the deployment region. Let M be a mobile node and K_s be the union of all the keys from the mobile key pool in the stationary nodes within the communication range of M. Let K_m be the number of keys from the mobile key pool S of size $|S|$ in M. The probability of establishing q secure links can be obtained by

$$P = \frac{\begin{pmatrix} |S| \\ q \end{pmatrix} \cdot \begin{pmatrix} |S| - q \\ K_m + K_s - q \end{pmatrix} \cdot \begin{pmatrix} K_m + K_s - q \\ K_m - q \end{pmatrix}}{\begin{pmatrix} |S| \\ K_m \end{pmatrix} \cdot \begin{pmatrix} |S| \\ K_s \end{pmatrix}} \tag{3.1}$$

In this equation we know the values of S,K_m and q. By fixing the value of P in the equation we can obtain the value of K_s. Key distribution in static nodes should be done so that the combination of all the static nodes in the neighborhood of a mobile node should have atleast K_s keys. Each node of the stationary network has e keys out of the key pool S of size $|S|$. The probability of a particular key from the key pool being in any node of the network is $\frac{e}{S}$. The probability of that key not being present in one node of the network is $\left(1 - \frac{e}{S}\right)$. The probability of that particular key being present in the x stationary nodes in the neighborhood of the mobile node is

$$P = 1 - \left(1 - \frac{e}{|S|}\right)^x \tag{3.2}$$

Therefore the total number of keys in the x neighbors of a mobile node K_m are

$$K_s = S.\left[1 - \left(1 - \frac{e}{|S|}\right)^x\right] \tag{3.3}$$

By fixing the value of P in the above equation, we obtain the value of x. This gives us the minimum number of stationary nodes within the communication range of the mobile node for it to establish a session key with probability P. If R is the communication range of the mobile and stationary nodes, then the total area in their communication range is πR^2 . For the mobile node to share q keys it needs x stationary nodes in its neighborhood. Based on this the required density of

stationary nodes d is

$$d = \frac{x}{\pi R^2} \tag{3.4}$$

The value d gives the minimum number of stationary nodes per unit area which would allow the mobile node to have q secure links from all points in the network with a probability P.

We now analyze the affect of node capture on the security of the scheme. Let c static nodes be captured. The capture of a static node compromises e keys. The probability of a key being compromised by the capture of a static node is $\frac{e}{|S|}$. The probability of a key not being compromised is $\left(1 - \frac{e}{|S|}\right)$. The probability of a key not being compromised after the capture of c static nodes is $\left(1 - \frac{e}{|S|}\right)^c$. Hence the probability of a key being compromised after the capture of c nodes is

$$P = 1 - \left(1 - \frac{e}{|S|}\right)^c \tag{3.5}$$

Equation 3.3 shows that an increase in the value of x would increase the value of K_s. Equations 3.1 and 3.5 show the tradeoff between security and connectivity. According to equation 3.1 an increase in K_s increases the probability of a mobile node sharing q keys with a stationary node in its neighborhood. On the other hand equation 3.5 shows that an increase in the value of $e\,(\propto K_s)$ increases the probability of a key being compromised incase of node capture. This gives us the tradeoff between security and connectivity.

Key Predistribution using Segmented Key Pool

In this scheme we have a large key pool which is divided into segments. Each of these segments is assigned to a sensor network. All stationary nodes randomly select keys from one of the segments whereas the mobile nodes select keys from the union of all these individual segments. Each mobile node randomly selects m keys from a key pool S. This key pool S is divided into n mutually disjoint segments

$S_1, S_2...S_n$. Each stationary node belonging to a network i obtains m keys from the segmentS_i. If a mobile node wants to establish q secure links with the network from any point of deployment and K_s is the union of all the keys in the nodes in the neighborhood of a mobile node. The probability of a mobile node with m keys and n segments establishing q secure links from a particular location in the sensor network is

$$P = \frac{\begin{pmatrix} |S_i| \\ q \end{pmatrix} \cdot \begin{pmatrix} |S_i| - q \\ K_s + \frac{m}{n} - q \end{pmatrix} \cdot \begin{pmatrix} K_s + \frac{m}{n} - q \\ K_s - q \end{pmatrix}}{\begin{pmatrix} |S_i| \\ K_s \end{pmatrix} \cdot \begin{pmatrix} |S_i| \\ \frac{m}{n} \end{pmatrix}} \tag{3.6}$$

By fixing the value of P,S_i,q,m and n in this equation we can obtain the value of K_s. Using the value of K_s and replacing e with m in equation (3)we can obtain the value of x which is the number of stationary nodes in the neighborhood of a mobile node which would allow the mobile node to establish q secure links with the stationary nodes with a probability P. Using the value of x in equation (4) we can obtain the density of node deployment.

The capture of nodes reveals the keys present in those nodes to the attackers. If the attacker captures c_i nodes from the network i then the probability of a key being compromised is

$$P = 1 - \left(1 - \frac{m}{|S_i|}\right)^{c_i} \tag{3.7}$$

An increase in the value of m would improve connectivity but worsen security. This tradeoff can be seen in equations 3.6 and 3.7. This is similar to the tradeoff seen between equations 3.1 and 3.5.

3.4.2 Simulations

In this section we analyze the performance of key predistribution using separate and segmented key pools. Our simulation considers a square deployment area of 200x200m^2 with the communication range of each stationary and mobile node being 20m. We assume all links to be symmetric meaning that if node A is within the communication range of node B then node B is in the communication range of node A. The capture of nodes by the adversary leads to the compromise of keys. In key predistribution using separate key pools the size of the mobile key pool is assumed to be 10000. In the segmented key pool based scheme the size of each segment is taken as 10000 and the number of keys in the mobile and stationary nodes is assumed to be the same. The number of keys in each mobile node are assumed to be 100. For clear understanding, these simulations we assume that the number of mobile nodes is equal to the number of stationary nodes although we believe that the number of mobile nodes would be much lesser.

In Figure 3.1 we show the relation between connectivity and the density of nodes. In this simulation we increase the number of nodes deployed and analyze the corresponding connectivity. Here connectivity is expressed as a fraction of links established to the total stationary nodes within the communication range of a mobile node. We calculate this value by placing the mobile node in 100 different locations of the deployment region. The increase in the number of nodes will increase the number of links formed because the mobile node can get connected to more nodes. But an increase in the number of nodes also means that the number of stationary neighbors to a mobile node increase. As a result the ratio of links established to the total neighbors is almost constant with the increase in stationary nodes. This figure also shows that key predistribution using segmented key pools has the best

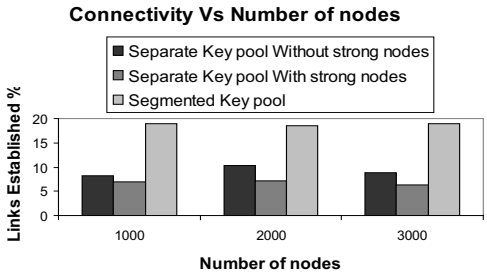


FIGURE 3.1. Relation between connectivity and the density of nodes. Here the memory overhead per node and the node capture are kept constant

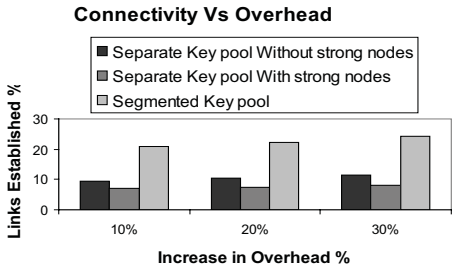


FIGURE 3.2. Relation between connectivity and overhead.In this case the number of nodes captured is kept constant

connectivity. This is due to the fact that a stationary node uses the same key pool to establish links with the mobile and stationary nodes.

Through these simulations we want to present the tradeoff's between security, connectivity and overhead. We plot graphs for all possible pairs of these values. Figure 3.2 shows the connectivity with respect to the overhead. For this simulation the value of node capture was kept constant. The increase in overhead results in better connectivity. This is expected because greater the number of keys stored, greater is the probability of the mobile node getting connected to the stationary nodes. Our simulations show that the the increase in the overhead is about the same as the increase in the number of links established.

In Figure 3.3 the relation between secure links compromised and the nodes captured is shown. As the number of nodes captured increases the attacker obtains more key information from the mobile pool and as a result more secure links

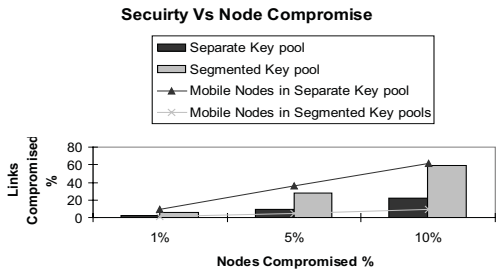


FIGURE 3.3. Relation between connectivity and security.In this case the memory overhead for the keys stored in the static nodes is kept constant.

are compromised. In case of key predistribution using a segmented key pool, the number of links compromised due to node compromise is very high because the same key pool is used by the stationary nodes to connect with other stationary and mobile nodes. For schemes with high rates of node capture, this scheme is would not be suitable. In the case of capture of mobile nodes, the segmented key pool scheme has an advantage because the number of keys from each segment of the key pool is small.

In Figure 3.4 we derive the relation between the increase in overhead and links compromised. We can see that as the overhead increases the links compromised also increase. An increase in overhead means that the number of keys stored in the nodes is increased. Although this results in better connectivity, the capture of one node would reveal a greater portion of the key pool to the adversary. As a result the capture of a node would compromise a lot more keys. We can see that when the nodes captured is kept constant the number of links compromised with the capture of each node increases with the overhead. In key predistribution using segmented key pools, the mobile nodes must store keys from all the different segments. Each segment is assigned to a different sensor network. As the number of different sensor networks which need to interact with the mobile node increases, the number of keys from the key pool of each segment goes down. This results in

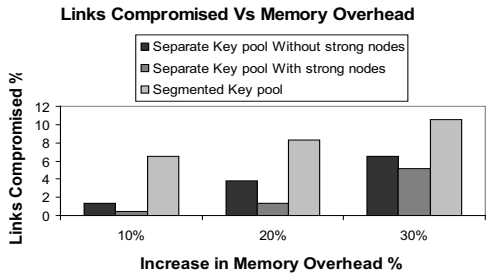


FIGURE 3.4. Relation between overhead and security. In this case the number of nodes captured is kept constant

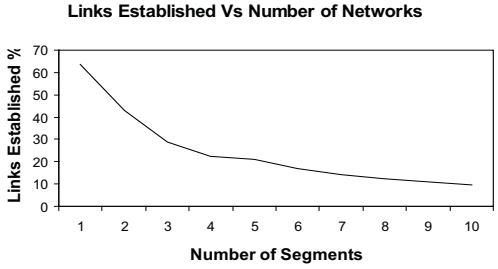


FIGURE 3.5. Relation between connectivity and the number of sensor networks. In this case the total keys in the mobile node is kept constant.

reduced connectivity between the mobile and stationary nodes of one particular sensor network. This trend is shown in Figure 3.5. The number of segments does not affect the links compromised because only stationary nodes are vulnerable to node capture.The number of sensor networks does not influence the number of keys compromised by the capture of each node.

3.4.3 Comparison of the Schemes

In this section we analyze the relative strengths and weaknesses of the key predistribution with separate key pools and key predistribution in segmented key pools.

In key predistribution with separate key pools, the number of keys stored in the stationary nodes is much less than in mobile nodes. The capture of stationary nodes leads to the compromise of a very small portion of the network. This scheme scales very well with the increase in sensor networks. The main disadvantage of this scheme is that the mobile key pool must be known before the deployment of

stationary nodes. The overhead of this scheme on the stationary nodes is due to the extra memory required to store the keys from the mobile key pool. This overhead is not there in the scheme using segmented key pools. Moreover the connectivity offered by using separate key pool for mobile nodes is less than that offered by the use of segmented key pools.

In key predistribution with segmented key pools, a large key pool is divided into disjoint segments and each of these segments is assigned to a sensor network. The stationary nodes randomly select keys from the key pool segment assigned to their sensor network and the mobile nodes randomly select keys from the whole key pool. This scheme allows the stationary nodes to communicate with other stationary and mobile nodes using the same set of keys stored in their memory. As a result this scheme avoids the overhead of storing extra keys unlike the schemes using a separate key pool. This also ensures better connectivity between the stationary and mobile nodes. The capture of a mobile node would compromise fewer keys between the mobile nodes and a particular stationary network. Also, unlike the previous schemes the keys compromised by the capture of stationary nodes in one network can not be used to compromise the links of another network and incase a network is extensively captured by the attacker, the mobile nodes can stop interacting with that network. The main disadvantage of this scheme is that it is not scalable if the number of networks becomes high. Although the use of one key pool means better connectivity, the number of links compromised incase of node capture is also much higher than the previous scheme. If the probability of the capture of the stationary nodes is higher the mobile nodes, the separate key pool scheme may be used. Otherwise the segmented key pool based scheme may be used provided that the number of networks in which the mobile nodes need to operate is low.

Chapter 4

Secure Emergency Communication of Cellular Phones in Ad Hoc Mode

4.1 Introduction

While electronic communications are the backbone of our civilization, they might make the difference between life and death for thousands if not millions of people during emergency conditions. Causes of such situations could be natural or man made disasters such as hurricanes, earthquakes, or terrorist attacks on infrastructure, biological attacks etc. An unfortunate example of an emergency situation and of the vital role of electronic communications was the aftermath of hurricane Katrina in south USA.

One of the most striking issues during and after hurricane Katrina was the loss of telecommunication infrastructure for both wireline and wireless networks. The lack of communications among authorities, first responders and population was the cause of consistent and unfortunate failures in directing and coordinating the rescue operation, with very tragic consequences.

While we have developed a very rich communication environment, including a very reliable telephone network, wired and wireless Internet, cell phones and WLANs, all of them fail when the respective infrastructure is damaged. We believe that the research community should develop technical solutions to guarantee necessary and vital communications in future emergency situations.

Due to their omnipresent nature, cellular phones (cell phones) have good a chance of becoming the best choice as the communication savior during emergency situations. For example, estimates that there are about two billions cell phones around the world with about 190 million only in USA [47]. However, today's cellu-

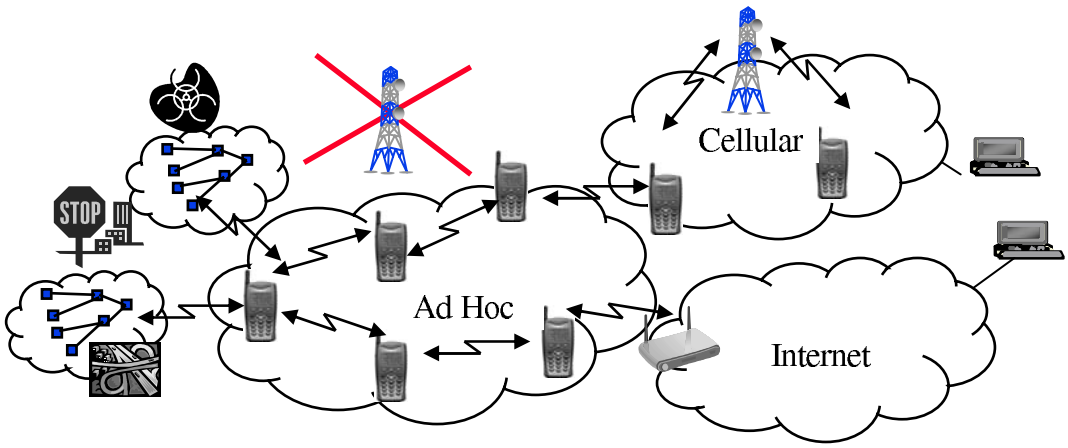


FIGURE 4.1. Ad-hoc network of cell phones. Initially, the base stations are alive and all cell phones are connected to the base station. When the base stations stop working, the effected cell phones communicate in ad hoc mode.

lar networks use fix infrastructures, which are vulnerable to the disasters' effects. As a consequence, in such conditions, while almost everyone might have a charged cell phone, he/she cannot communicate because his/her base stations are damaged [10]. One good alternative in such conditions is to switch and use cell phones in ad hoc mode.

The communication architecture that we assume for emergency situations, is illustrated in Fig. 4.1. We assume that cell phones, besides their normal cellular interface, are equipped with physical interface such as the 802.11 family to communicate among themselves in ad hoc mode when their infrastructure is no longer available. We also assume that, in emergency conditions, broadcast will be the most needed model of communication. In addition, we also assume that cell phones can connect to both the base station and the ad-hoc network at the same time as shown in Fig. 4.1. The ability to change from centralized mode to ad hoc mode will allow cell phones to leverage the advantages of both systems.

Ad hoc networks consist of self configurable mobile nodes which communicate through the wireless medium. These networks do not require a centralized base

station and nodes may communicate through multi-hop routing. The nodes are free to move randomly which makes the topology highly dynamic and unpredictable[60]. Research work has been done related to how to integrate cellular networks and WLAN [38],[4]. Also, significant research efforts are dedicated to ad hoc networking [45], [28]. However, these works do not consider conditions and restrictions created by emergency and disaster situations. For example, in disaster conditions energy saving becomes an important goal, as it may be impossible to charge cell phones, which makes them in some aspect similar to sensor networks in some others to ad hoc networks [5, 6, 49], while showing some unique characteristics. We list some characteristics of cell phones in ad hoc mode that are crucial for the design and development of security solutions.

- Cell phones have significantly more computational power, memory and battery life than the resource starved nodes of sensor networks. However, if emergency conditions persist, conserving energy might become important for cell phones in ad hoc mode. On the other hand, the importance of the information to be transmitted/forwarded/received could be such that there is no sense to save energy.

- While the communication range of cell phones is much larger than that of sensor and most of classical ad hoc, transmitting at maximal range will drain cell phones' batteries.

- The major novel characteristic of cell phones in ad hoc mode, compared to both sensor and classical ad hoc networks, is that the phones were once part of a secure centralized network. When designing the security of cell phones in ad hoc mode we should take advantage of this unique characteristic. For

example, a cell phone can obtain the latest key information when it is part of the centralized network and use it when the phones are in ad-hoc mode.

The uniqueness of cell phones in ad hoc mode is strengthened even more by the nature of the traffic they have to carry. Some examples of such traffic could be: urgent messages sent by authorities to warn the population about particular situations in the city, subways, buildings and other infrastructures; information gathered in various ways including sensor networks (shown in 4.1) could be sent to all interested people regarding pericles to be avoided, unique ways to escape from life threatening situations. As shown in these examples, most of the communications in emergency conditions are broadcasts. Therefore, we assume that some efficient existing or future broadcast protocol will be used.

The broadcast applications in emergency conditions require the data to be transmitted to all the users; hence, data confidentiality is not a requirement. We assume two major models of security attacks. In emergency conditions, the population is very vulnerable to malicious information, which could go from pranks by irresponsible individuals to the use of such malicious information as important part of the attack by terrorists. An example of the second case could be messages directing people towards specific infected area in order to maximize the effects of a biological attack. Therefore, the primary requirement for such applications is broadcast authentication with guaranteed non-repudiation, which would lead to identification of sources of malicious information. While this protection might be enough to discourage irresponsible individuals, it will not stop terrorists who are even ready to commit suicide to reach their goals. In this case it is very important to discover the malicious nature of the information as soon as possible and warn the popula-

tion about the it as well as the source itself. The goal of our paper is to propose solutions to secure the broadcast communication from the above described attacks.

Security in an ad-hoc network of cellular phones is very different from that in sensor and classical ad hoc networks[22, 13]. The capture of cell phones is trivial because the attacher could simply buy them. In case a cell phone is lost, it is assumed that the owner revokes it. Also, the communication between the base station and the cell phones is always assumed to be secure.

For the secure authentication in cell phones in ad hoc networks we propose to use Identity based cryptography [8, 27, 51] In Identity based Cryptography the ID assigned to a mobile phone by the base station is used as the public key. The private key for the corresponding public keys (ID's of the cell phones) are generated by the Private Key Generator (PKG) in the base station. The base station can assign ID's and private keys when it authenticates the mobile phones. The most recent key information given to the phones by the base station is used when they are unable to connect to the base station.This approach restricts the mobility of the phone to within the cell in the ad hoc mode. To allow a phone to move in the ad hoc mode, the base stations ensure that the each phone is able to communicate securely with a fraction of the phones in its neighborhood.

To protect the population from malicious information, we assume that a fraction of the cell phone users are able to detect at some point in time the real nature of the malicious message. How an individual reaches such decision is out the scope of this paper, but examples could be contradictory information, personal verification of the information, etc. In our scheme every cell phone that decides that the received message is malicious broadcasts the identity of the sender to warn other mobile phones. The distributed nature of this scheme prevents attackers from discrediting

genuine phones. We assign trust levels to each cell phone and use this level of trust in deciding if the message is malicious or not.

The structure of this paper is as follows. We discuss the existing literature related to this paper in Section 4.2. We present the scheme for non-refutable secure broadcasts using identity based cryptography along with the approach for detection and revocation of malicious nodes in Section 4.3. Analysis for these schemes is included in Section 4.4.

4.2 Background Work

In 2-2.5 G cell phone systems such as GSM, each cell phone has a Universal Integrated Circuit Card (UICC), also referred to as SIM card[37, 41]. It acts as the ID of the phone and stores all the information related to the working of the phone. To authenticate a mobile phone the base station sends a 128-bit random challenge ($RAND$). Each phone has a 128-bit Individual Subscriber Authentication Key ($ISAI$) stored in the SIM. Using the $RAND$ and $ISAI$ the mobile phone generates a response to the challenge posed by the base station along with the session key for the communication. This session key is used to encrypt the communication between the mobile station and the base station. In this paper we propose to assign a new ID and the corresponding private key to the mobile phone when it is authenticated.

Identity based cryptography was first proposed by Shamir in 1984 [51]. In this paradigm, the users' ID, like phone number or email ID, has a one-on-one mapping with the public key of the user. This reduces the system complexity and cost for establishing and managing public key infrastructure [8, 27]. The Private Key Generator (PKG) creates its master and public key. A user can get authenticated to the PKG and obtain the private key corresponding to its ID. Any other user

can obtain the public key of the user using his ID and the public key of the PKG. The biggest advantage of Identity based cryptography over traditional public key cryptography is that there is no requirement for looking up public keys, and one of the big practical difficulties that has been associated with public-key cryptography is no longer an issue. Being able to calculate public keys is particularly useful in situations in which an entity needs to communicate securely with an unknown party. Identity based Cryptography also allows a user to communicate that an entity which is not enrolled in the system.

4.3 Authenticated Broadcast with Non-Repudiation

4.3.1 Mechanism

Our goal is to provide non-refutable authenticated broadcast of cell phones, which switch to ad hoc mode when they are unable to connect to the base station. The level of resources, such as memory and energy, allows us to propose asymmetric cryptography in case of emergency situations when cell phones are in ad hoc mode. Although symmetric key cryptography consumes much less energy than public key cryptography, achieving source authentication is difficult because each mobile phone has the same information. To avoid the overhead of verifying certificates every time, we make use of Identity based Cryptography. In Identity based Cryptography, a publicly known identity of a user can be used to derive the public key of the user.

Our scheme assumes that each cell of the cellular network has a Private Key Generator(PKG) which is trusted by all cell phones in the cell. The PKG generates a master key and some system parameters to enable Identity based Cryptography. The master is kept private by the PKG whereas the system parameters are public. For clear illustration we refer to master key as private key(SK_{pkg}) of the PKG

and the system parameters as the public key(PK_{pkg}) of the PKG. Each mobile phone is given the public key of the PKG. PK_{pkg} is used to generate the public keys of a mobile phone using its ID. Using the PKG, the base station sends each phone in its cell, an ID, the corresponding private key along with its public key. A disadvantage with Identity based Cryptography is that it needs a third party in the form of PKG, which knows about the secret key. Therefore, non-repudiation is difficult to prove. Moreover, the need for a secure channel between the users and the PKG even before channel security is established is also another major disadvantage. These disadvantages are taken care of in cellular networks where the base station has a secure communication channel with all cell phones. Securing the base stations is important whether the cell phones are allowed to communicate in the ad hoc mode or not. To minimize the damage caused by the compromise of base stations we use multiple PKG's in a cellular network.

Every mobile phone has a 128-bit Individual Subscriber Authentication Key ($ISAI$) stored in the SIM. When the phone is turned on, the base station sends a random challenge ($RAND$) to the cell phone. The phone calculates the response to the challenge and sends the response to the base station. If response received by the base station matches the calculated response, the cell phone is authenticated. The cell phone is able to calculate a session key K_s using $RAND$ and $ISAI$. The base station sends each mobile phone an ID and the private key corresponding to the ID. This message is secured by using the session key K_s. In the ad hoc mode when a cell phone wants to broadcast a message, it signs the message with private key and floods the network with this message using an efficient broadcast schemes such as [20]. Any receiver of this message which has the public key of the PKG of the sender may obtain the public key of the sender and verify his signature. This

scheme used for authenticated broadcasts and does not provide any confidentiality because all mobile phones can generate the public key of mobile phones.

Let BS represent a Base station and CP represent a Cell Phone. We show the sequence of communications between the base station and the mobile phone.

$$BS \rightarrow CP : RAND \tag{4.1}$$

$$CP \rightarrow BS : RESP = f_1\left(RAND, ISAI\right) \tag{4.2}$$

$$CP : K_s = f_2\left(RAND, ISAI\right) \tag{4.3}$$

$$BS \rightarrow CP : K_s\left[ID, SK_{ID}, PK_{pkg}\right] \tag{4.4}$$

The function f_1 implements the $A3$ algorithm which generates the 32-bit signed response using the 128-bit $RAND$ and the 128-bit $ISAI$. This response is used to authenticate the mobile phone to the base station. The cell phone then generates the session key K_s using function f_2, which implements the $A8$ algorithm. After this stage the base station can establish a secure communication with the mobile phone using the session key K_s. After the security is set up for the cellular network, the base station can distribute the information for secure communication in the ad hoc mode. The base station generates a private key for each cell phone corresponding to its ID and sends it to the mobile phone. This message is encrypted using the session key K_s. The mobile station generates its public key PK_ID using the public keys of the PKG. To send a message, a cell phone signs the message using its private key. The receiver of this message calculates the public key of the sender using the ID of the sender and the public key of the PKG. This is clearly shown in equation 4.7.

$$Sender : C = SK_{ID}\left[M, ID\right] || ID \tag{4.5}$$

$$Receiver : M = PK_{ID}\,[C] \tag{4.6}$$

Here the public key of the cell

$$PK_{ID} = f\,(ID, PK_{pkg}) \tag{4.7}$$

Since a cell phone can not obtain the private key of another phone, this scheme ensures non-repudiation in the ad hoc mode.

4.3.2 Key Revocation

Unlike the addition of a user to the network, the revocation of a user is much harder. When a user decides to end his subscription with the service provider or when a cell phone is lost, the cell phone must not allowed to make use of the key information stored in the phone to connect to the network. In the normal mode, the base station has all the information of each node, which makes revocation easy. This is unlike the ad hoc mode, in which the mobile node has no information about which nodes have been revoked. A naive solution to this problem would be store the ID's of all phones which have been revoked in each active cell phone of the cellular network. Although simple, this solution has the disadvantage that for a large number of cell phones distributing this information is cumbersome especially in the ad hoc mode.

A possible solution to this problem is to include timestamps in the IDs assigned to each cell phone. In such a case the ID assigned to a cell phone and the corresponding private key sent to the node by the base station is valid only temporarily (until the timestamp expires). These temporary IDs and keys need to be refreshed regularly as long as the cell phones are connected to the base station. The base station can stop refreshing the keys of cell phones that have to be revoked. When the base station fails, these nodes would use the latest ID and key assigned to them. In addition, when a base station fails, all the other base stations transmit

this information to their cell phones. This enables cell phones to communicate with phones that have old IDs and whose base stations have collapsed.

4.3.3 Secure Mobility Management

Mobility is fundamental to cell phones. In this section we discuss the problem of secure mobility with multiple PKG's and present a solution which provides a tradeoff between various communication overheads, discussed in more detail in Section 4.4.1. The PKG in the base station of every cell assigns IDs and private keys to all the cell phones in that cell. All the phones in a cell have the public key of the PKG, which they use to generate the public keys of other mobile phones. A phone p of cell A has an ID and private key assigned to it by the PKG of cell A. Moreover, it also has the public key of the PKG of cell A with which it verifies the signatures of the other mobile phones in cell A. When a cell phone moves to another cell in the ad hoc mode the situation is different. If the phone p moves to a cell B, other cell phones of B would not be able to verify the signature of p because the private key of phone p has been assigned by the PKG of cell A. Similarly, phone p would not be able to verify the signatures of other phones of cell B because it does not have the public key of the PKG of cell B.

Any two cells A and B are n^{th} degree neighbors if there are $(n-1)$ cells between them. For example, any two neighboring cells are first degree neighbors of each other. To establish seamless mobility for a phone in its k^{th} degree neighborhood, the base station of the cell broadcasts the public keys of the PKG's of all the cells in its k^{th} degree. Since k is a network wide parameter decided by the network administrators, if phones of cell A have the public key of the PKG of cell B, then phones of cell B will have the public key of the PKG of cell A. When phone p of cell A goes to cell B, it can verify the signatures of the phones of cell B using the

public key of the PKG of cell B, whereas the signature of phone p can be verified by the phones in cell B using the public key of the PKG of cell A.

If the two phones are not in the k^{th} neighborhood of each other, then establishing communication between them is non-trivial. When a phone p of cell P moves into a cell Q which is not in the k^{th} neighborhood of P, it can neither send nor receive the broadcasts in cell Q. Phone p can simply request for the public key of the PKG of cell Q from any phone of cell Q. Using this public key, the phone can read all the broadcast messages of the cell. For the phone p to send broadcasts in cell Q, all phones in cell Q should have the public key of cell P. A geographic routing protocol such as the Geographic Energy Aware Routing(GEAR) [59] can be used to query the public key of the cell P. In this scheme, upon receiving a packet the nodes would forward it to a neighbor closer to the target region than itself. Once the packet reaches the target region, it is diffused through techniques such as flooding. The reply for this packet is sent using the same scheme in the opposite direction. This reply is signed by the cell P using the public key of its PKG. Whenever the reply moves into a different cell, the signature on the reply from the previous cell is removed and a new signature is added. As a result, cell Q would receive the public key of the PKG of cell P signed by a cell in its first degree neighborhood.

4.3.4 Detection and Revocation of Malicious Phones

The scheme discussed thus far offers non-refutable broadcasts for cell phones in ad hoc mode. In case a cell phone transmits malicious data, the non-refutable nature of these broadcasts ensures that the malicious users can be detected after the base station is reactivated using the public key of the malicious phone. Although the detection of the malicious phone is guaranteed, the time taken for the base

station to get reactivated may be considerably long. This gives the malicious phones considerable time to transmit malefic messages without being detected or revoked. Such vulnerabilities could be exploited by terrorists to spread false information within the cell phones and maximize the damage of their terrorist attacks.

The problem becomes more complicated when the malicious phone is mobile and its spreads malefic messages throughout the network. The goal of our scheme is to minimize the damage caused by the malicious phones by spreading the revocation information in the entire network in the ad hoc mode. Preventing the malicious phones from discrediting non-malicious phones (which is a form of Denial of Service) also becomes challenging in large networks consisting many cells because the revocation information has to be spread throughout the network.

There are two aspects in the detection and revocation of malicious nodes. The first part is the decision making process in which cell phones collaboratively decide if a particular phone is malicious or not. The second part of this problem is to communicate this decision throughout the network in such a way that it prevents malicious phones from revoking other non-malicious phones. We discuss these two stages.

- **Decision Making:** When a cell phone receives a message which it believes is malicious, it broadcasts a local revocation message for the malicious phone to all its neighboring phones. Each phone maintains a counter for the local revocation messages against a phone. To prevent the faking of these messages, they are signed by the private key of the sender. The counter of each phone is independent of the other phones. When the counter reaches a threshold value, a network revocation message is broadcasted throughout the network. The decision making process is collaborative and independent i.e. although

all the cell phones would collaborate to make the decision, the decision of each phone is independent.

- **Decision Broadcast:** When the revocation counter value in a cell phone crosses the threshold, the phone broadcasts a network revocation message against the malicious phone using efficient broadcast protocols such as BPS [20]. To prevent malicious phones from revoking other non-malicious phones, this network wide broadcast includes all the local revocation messages based on which the network revocation message is being sent. When a phone receives the network revocation message, it verifies the signatures of all the local revocation messages. If all the signatures are correct, it forwards the message by removing the local revocation messages and signs the revocation message with its own signature. This forward message is also received and verified by the previous sender of the message. If a phone forwards a message other than the one that it received, its neighboring nodes would detect that the sender is a malicious node and broadcast this through out the network. This scheme ensures that the revocation message is broadcasted throughout the network.

This is clearly illustrated in equation (4.8) in which $S_1, S_2 \ldots S_n$ are the local revocation messages and RM is the network revocation message. Also let P_1 be the first phone whose local revocation message counter exceeds the predefined threshold. Let P_2 be the node which forwards the broadcast message.

$$\begin{aligned} P_1 \rightarrow Neighbors : SK_{P_1}\left[RM, S_1, S_2 \ldots S_n\right] || P_1 \\ P_2 \rightarrow Neighbors : SK_{P_2}\left[RM\right] || P_2 \end{aligned} \tag{4.8}$$

If the phone P_1 is malicious, then it will not be able to produce the local revocation messages $S_1, S_2 \ldots S_n$. It is assumed that the number of malicious phones in a locality is less than n. The selection of the threshold n also presents a tradeoff. If the value of n is too small, the probability of having n malicious phones in the neighborhood is high, whereas a higher value of n would mean greater transmission overhead and computation time. If the phone P_2 is malicious than it would not repeat the network revocation message sent by P_1. In that case P_1 would know that P_2 is malicious and start the revocation process for P_2.

4.4 Analysis of Secure Broadcast Scheme

In this section we analyze the tradeoffs in the scheme and how the choice of certain parameters would affect the performance of the scheme.

4.4.1 Value of k

The phones in each cell are assigned keys of all the PKG's in the k^{th} degree neighborhood of the cell. The value of k is a design parameter decided by the network administrators. If the value of k increases, the key of the PKG of a base station has to be broadcasted (while in cellular mode) to more cells and more public keys of cells would have to be stored in each phone.

If the value of k is low and a phone moves beyond the k^{th} degree neighborhood, there will be a lot of communication overhead among cell phones in ad hoc mode to obtain the public key of the PKG which has assigned the private key to the phone in motion. The advantage of a higher value of k is that the phone can move seamlessly in a greater portion of the network. If the scheme assumes that the period for which the base station would be dysfunctional is small, then the overhead caused by k might not be justified. On the other hand, if the base station is likely to be out of order for long periods of time and the phones are likely to move

beyond the k^{th} degree neighborhood, then a higher value of k may be justified. A higher value of k would result in the overhead on base stations (on cellular mode) increasing because they have to broadcast the public keys of the PKG's of more cells. On the other hand, a higher value of k would reduce the communication overhead on each phone (in ad hoc mode).

In Figure 4.2, the results were obtained by placing a cell phone outside the k^{th} degree neighborhood of the cell to which it belongs. The results are given with the transmission range r at different ratios to the maximum range R. In this analysis we calculate the overhead in the cellular and ad hoc mode by placing a cell phone in its tenth degree neighboring cell. If the overhead message in the cellular mode is 1, then the public key of the PKG of the cell phone in question is broadcasted in only one cell. When the communication range of the phone is same as the radius of the cell, the number of broadcast messages required for a tenth degree neighbor to obtain the public key of the PKG of the cell to which the cell phone belongs is 18 (9 for sending the query and 9 for receiving the reply). All the other points on Figure 4.2 have been obtained simulating using this procedure. Figure 4.2 clearly shows that a decrease in the overhead of individual phones results in an increase in the overhead messages of the base stations.

4.4.2 Threshold for Node Revocation

Revocation of malicious cell phones in the ad hoc mode is very critical in emergency conditions. A single bad node could broadcast data throughout the ad hoc network and cause widespread misinformation. Whenever a phone detects that some information being broadcasted in the network is malefic, it will send a local revocation message to all the phones in its communication range. Each phone aggregates the number of revocation messages against a sender and if this number

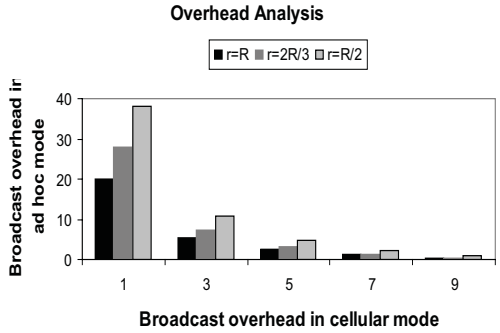


FIGURE 4.2. Tradeoff, depending on the value of k between base station overhead messages (in cellular mode) and cell phone overhead messages (in ad hoc mode). The results are given for various rations between actual (r) and maximum (R) transmission ranges.

exceeds a Revocation Threshold value Th_r, a network wide revocation message is sent against the sender.

This threshold value Th_r is an important network parameter. If the threshold is too low, a malicious user could discredit the genuine mobile phones of the network. This is a form of Denial of Service. A higher value of threshold might prevent a malicious phone from causing a Denial of Service attack on the network because the number of malicious phones required would be higher. We make the assumption that the value of the threshold is proportional to the time taken to send the network revocation message. If the threshold Th_r is high, the time taken to reach the threshold would be high, thereby giving the malicious phone, more time to spread the malefic information. On the other hand if the threshold is low, it can be reached quickly and the damage caused by the malicious phones can be reduced.

In Figure 4.3, the Contribution to Denial of Service (DoS) that a malicious phone makes is captured. If the threshold Th_r is one, then the contribution to the DoS made by a malicious phone is one, meaning that is enough one node message to consider another node as malicious. For a threshold Th_r value of two, the contribution would be 0.5 which means that two malicious phones are required

to discredit a genuine phone. As expected, the increase in the threshold value increases the number of malicious phones required to perform DoS. To quantify the effect of spreading misinformation we use the metric "Damage", which is equal to the number of cell phones that receive the malefic message before a network revocation message is sent. Figure 4.3 shows the Damage relation to the threshold (time). We assume, in a first approximation, that this value is directly proportional to the number of cells which receive malefic information before the revocation message is sent. Suppose one unit of time is taken to broadcast information in a cell. In a simple hexagonal structure of equal cells, after the second unit of time six new cells surrounding the central cell have malicious information which makes the total cells with malefic information equal to 7. After the third unit of time the damage increases to 19 and so on. Depending on the structure of cellular network and the number of phones in each cell the Damage increases exponentially with time.

The revocation of phones depends on the perception of users of the network. There may be other factors which influence human decisions. For e.g. in the case of Denial of Service in Figure 4.3, when some phones say that a particular phone is malicious, the human tendency would be "to go with the flow". The curve for Denial of Service produced after considering these factors may differ from the line behavior shown in Figure 4.3. We believe that these factors need to be carefully analyzed and that they are beyond the scope of this work.

4.4.3 Analysis for Base Station

The key information for the ad hoc mode is provided to the nodes by the base station using the secure cellular infrastructure. The implementation of this scheme requires the base station to update the key material at regular intervals for better

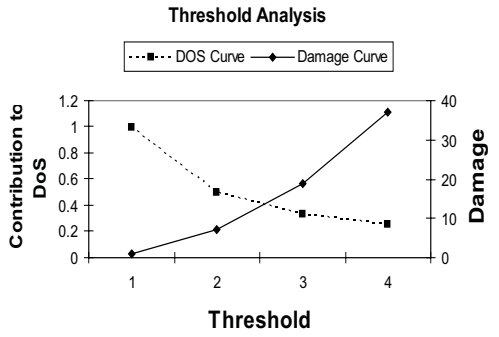


FIGURE 4.3. Tradeoff between the Contribution to Denial of Service and the potential Damage to the network. This tradeoff depends on the Revocation Threshold Th_r

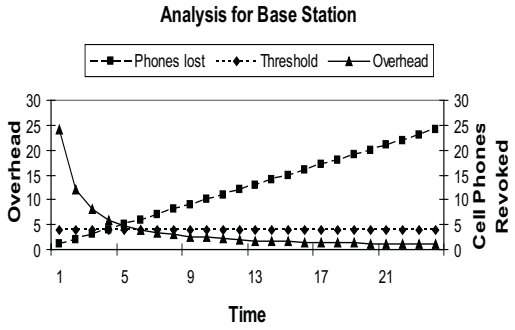


FIGURE 4.4. This figure presents the relation between overhead for the base station and the time duration for key refresh. It also shows the rate at which cell phones are lost and the threshold for the number of cell phones lost after which keys have to be refreshed

security and managing node revocation. The frequency at which the information is refreshed is an important system design parameter for the network administrator. If the information is refreshed very frequently, the overhead on the base station for generating and transmitting the new keys is very high. It also improves the security of the scheme because the cell phones which have been lost or stolen are revoked quickly. On the other hand if the time duration between the refresh of keys is high, the overhead on the network is low and the security of the scheme decreases. The relation between time for key refresh and security is shown in Figure 4.4. The frequency of key refresh is based on the rate at which cell phones are lost or stolen. When this value crosses a threshold, the keys have to be refreshed to

revoke the lost cell phones. Figure 4.4 clearly shows that as refresh time increases, the overhead decreases. The number of phones lost is a linear curve.

Chapter 5

Reputation Based Revocation in a Ad hoc Network of Cell Phones

5.1 Introduction

There are many applications which would greatly benefit from using cell phones in the ad hoc mode. In many natural disasters like hurricanes and earthquakes where the cellular phone and the traditional phone infrastructure is damaged. Enabling the cell phones to communicate in the ad hoc mode provides mechanisms for people trapped in buildings or elevators to seek help. These networks can also help the agencies plan and coordinate the relief effort. These applications require data to be transmitted to all users hence data confidentiality is not a requirement. The ease with which a user can communicate with a large number of other users makes this system vulnerable to attacks ranging from pranks to terrorist attacks. Terrorists could artificially increase the population of an area by asking all other cell phone user to go there before a terrorist attack. To ensure that the perpetrator of these attacks is identified, these schemes require secure broadcast authentication to achieve guaranteed non-repudiation. This approach is reactive because it identifies the attacker after the attack. Clearly this is not enough for some attacks that may are possible with this system. Protocols need to be developed which would prevent a malicious cell phone user from misguiding other users[9].

In this chapter we present a scheme for reputation management and node revocation in a distributed ad hoc network of cell phones. This is an extension of the work presented in Section 4.3.4. Each phone would have a reputation score which is uniquely mapped to its ID. From here on we refer to the cell phone as node and the reputation score as score. When a node believes that another node is malicious

it attempts to reduce the score of the node perceived to be misbehaving. When the score of a node becomes less than a predefined threshold, the node would then be revoked. This scheme also takes into account the fact that a malicious node may attempt to revoke a honest node.

When the nodes are connected to the base station, they can obtain the necessary key information which is used to communicate securely in the ad hoc mode. The base station assigns keys and ID's for the nodes in the ad hoc mode. The Id's in the ad hoc mode are not related to the ID's in the normal mode to ensure anonymity. For each node, there are n other nodes which are given a share of the key required to sign the score certificate. Out of the node certifiers k need to agree on a new score and sign a new certificate. When the network moves into the distributed mode from the centralized mode, each node has to obtain certificates of score from its anonymous certifiers at regular intervals. Whenever a receiver of a broadcast message believes that the sender is malicious, a message is sent to the anonymous certifiers of the node. The certifiers remain anonymous because nodes communicate with them in the ad hoc mode without any knowledge of their real identities. When the number of these messages becomes large, the anonymous certifiers give the node a new score certificate. There may be special nodes in the network which have more than one share of the group key. These trusted nodes may be important people like cops, mayors etc who we believe can be trusted. The number of shares of the key given to each of those trusted nodes is a design parameter. We believe that the efficient use of these trusted nodes would give us the same level of security with lesser overhead.

Reputation is defined as the perception that a node creates through past actions about its intentions and norms [43]. The reputation of an node is based on the perception regarding its behavior held by other agents based on its experiences

and observations of its past actions [36]. Such experiences are conveyed through recommendations which can be both positive and negative. The major challenge for reputation based systems is assessing the truthfulness of such recommendations. In ad hoc networks, the absence of a commonly trusted entity means that the reputation system has to be distributed to the nodes of the ad hoc network. When the system is distributed, there may be an issue of recommendations that may contradict each other. The reputation systems in ad hoc networks need to be robust against contradicting reputations[44].

The three main attacks against a reputation based systems are the free rider problem, defamation and collusion [36]. In the free rider problem nodes do not share the reputation information with their peers. If a significant number of nodes don't communicate about the misbehavior of malicious nodes, then the nodes of the network will not be able to identify the malicious nodes. We believe that this problem can be solved by rewarding people for sending information about malicious nodes. Although this is an important problem, its solution is out of the scope of this paper. The second attack on a reputation based system is defamation. In this attack a malicious node attempts to defame a honest node and tries to get it revoked. Our scheme handles this problem by increasing the number of nodes which believe that a particular node is malicious. Even if a malicious node is part of the anonymous certifiers, it can't sign a revocation certificate without $k-1$ other nodes. The third common attack on repuatation based systems is collusion. Here a node tries to improve the score of another node by sending multiple positive feedbacks. Our scheme avoids this problem because scores are not increased in the ad hoc mode. They are increased once the nodes go back to the centralized mode. The central base station would be able to track multiple recommendations from the same node and prevent collusion.

This paper provides a mechanism to adjust the reputation score of a node in the ad hoc mode using threshold cryptography. The scheme also allows a group of anonymous certifiers to revoke a node based on the recommendations of other nodes. These recommendations are also subjected to scrutiny and there may be recommendations which provide feedback on the quality of recommendations. This would ensure that nodes which broadcast malicious recommendations are also revoked. The recommendation messages are designed in such a way that the recommenders and the recommended can be identified unambiguously. All the information can be collected once the base station is back online. The base station can then determine the malicious nodes. We would like to mention that the recommendations made by the users of the network are based on their own experiences which may not be similar to the experiences of other nodes. This paper doesn't deal with the aspect of decision making in these networks. There are many factors which may influence decision making which are out of the scope of this paper.

The structure of the paper is as follows. In section 5.2 we discuss the system model for which we propose our scheme. Section 5.3 has some related work in the area of trust and cryptography. Section 5.4 describes our scheme in detail. Section 5.5 presents the theoritical analysis and simulations.

5.2 Related Work

This chapter uses the protocol for establishing non-refutable broadcasts in an ad hoc network of cell phones presented in Chapter 4. These networks have the special property unlike other ad hoc networks, which is that the nodes of the network are part of the centralized network before they enter the ad hoc mode. This allows these base station to send all the key information securely using the secure cellular infrastructure. We achieve secure non-repudiation in these networks by the use

of Identity based cryptography. Details on the implemenation of this scheme are presented in the next section.

Kinateder and Rothermel describe two usage scenarios for a reputation based system, namely publishing recommendations and requesting recommendations [31]. In a publication based system, an entity which interacts with the target and decides to create a recommendation, publishes(broadcasts) the recommendation to all nodes of the network. On the other hand a requesting recommendation schenario is one where the entity which has to interact with the target enquires about the reputation of the target. In this paper, we present another usage scenario where, the node in question stores its own reputation value which is signed by the anonymous certifiers.

Many of the desired properties in a reputation based scheme for mobile ad hoc networks have been listed in [36]. The system should be able to unambiguously distinguish between the malicious and non-malicious nodes without the use of any centralized infrastructure. The system should also be robust to common attacks like node inactivity, defamation and collusion by a small number of nodes. Another important requirement is the timeliness of the information provided. The system would not useful if the malicious nodes are detected long after all the damage is done. Our scheme provides a trade-off between the communication overhead and the timeliness of reputation information.

The lack of infrastructure in the ad hoc mode means that the schemes for trust establishment can only depend on the local interaction of the nodes. In this situation trust management should start with a small group of trusted nodes which gradually establishes trust with the initially neutral members of the network. The whole network evolves from trust islands into a trust graph [56]. In the context of cell phones in the ad hoc mode, a broadcast would be trusted if the node can ob-

tain a certificate from one of the few trusted nodes in the network (like cops/mayor etc). If the node leaves the neighborhood of the trusted nodes, then their broadcasts have no trust. This approach may not be scalable if the number of nodes is much greater than the number of trusted nodes. Moreover, mobility of various nodes would also cause problems with location dependent security.

[60] utilizes threshold cryptography for the distribution of trust in the ad hoc mode. In this scheme signing key of the CA can be split into n pieces such that each node gets a piece of the information. Any k of these n can be used to recreate the private key of the CA. Even if $(k-1)$ malicious nodes collaborate, they would not be able to generate the key information. The disadvantage of this scheme is that if the number of adversaries becomes more than k the security of the entire network is compromised. In our scheme, security decreases much more gracefully with the increase in the number of attackers.

[15] presents a scheme which assumes that all nodes are connected to each other all the time. They all maintain a profile table which has the accusations against all the nodes of the network. The weights of the accusations of all nodes are different. The weight of the accusation of a node depends on the number of accusations against itself. This paper assumes that the all the nodes of the network would have exactly identical profile tables and those which have a different profile table are malicious.

5.3 System Model

The nodes of the network are cell phones. They have significant battery power, computational power, communication range and memory. This makes them suitable for the use of asymmetric key cryptography. Our scheme is useful only when the cellular phone infrastructure is dysfunctional. Reputation Management and

node revocation are trivial in a centralized system because all nodes can communicate only through the base station. We assume that the number of base stations that would be impaired at any point of time would be small. This combined with the fact that the communication range of nodes is large, means that a node can send data to the entire network in few hops. Also we assume that the density of nodes in the deployment region is large. It is also assumed that more than k of the n anonymous certifiers for each node would be honest. The values of k and n are calculated after analyzing the tradeoff between security and overhead.

The base station of the cellular network distributes all the key information which would be used by the nodes when they are unable to connect to the base station. The base station creates an ID for each node which can't be derived from the ID of the phone. We believe that Identity based Cryptography(ID-PKC) [8, 51] would be best suited to this scheme. The base station generates a private key corresponding to each ID using some private information. The private key of each node is securely transmitted to the node using the cellular infrastructure. ID-PKC avoids the overhead for the verification of certificates because the public keys are generated using the ID's. If each node has to generate a new public/private key pair and get the public key signed by the CA (base station), the overhead is high. Instead if the base station generates the $< ID, PrivateKey >$ for each phone and sends this pair to the phone using the secured cellular infrastructure, the overhead is considerably less. This scheme also takes the burden of key generation from the cell phone and gives it to the base station. The private keys corresponding to an ID are generated using a Private Key Generator(PKG) of the cellular network. Any node with the public key of the PKG can derive the public key of a node using its ID.

An Ad hoc network of cell phones has more vulnerabilities than other instances of the mobile ad hoc networks because node capture is trivial. One can not prevent attackers from being parts of the network because cell phones are freely available in the market for purchase. Any security scheme proposed for this environment must be able to address this problem. In this paper we assume that the number of malicious nodes in the network is very small compared to the total nodes of the network.

Based on the number of the maximum number of malicious nodes expected, the tradeoff for the level of score when a node is revoked are decided. We define the reputation score from 0 to 1, where 0 signifies no trust and 1 signifies maximum trust. The initial score for each node is assigned by the base station. This score may be based on many factors like criminal history, credit history etc which are out of the scope of this paper.

The honest user of the network is the most important part of the detection of the malicious nodes. Whenever a cell phone of a user receives a message, the user of the cell phone analyzes the message and performs the appropriate actions for his safety. But, when a user believes that the message was incorrect(or malicious) he sends a negative recommendation to all the anonymous certifiers of the sender. Free rider problem is a major problem in reputation based systems. This deals with the willingness of the attacker to report the malicious behavior of a node. In this paper we assume that a honest node would report the malicious behavior of a node because of the horrific consequences which may be caused if the user doesn't respond. We believe that a combination of laws with punishments for withholding information and reward for honest participation may also help. This aspect of the protocol is for the law enforcement agencies to address and we belive that it is out of the scope of this paper[25].

There may also be a incentive based approach where the honest user gets some benefits for reporting the malicious users.

5.4 Our Scheme

In emergency situations when the base stations of the cellular network become dysfunctional, the nodes collaborate and form an ad hoc network. These networks can be used by users to communicate information and coordinating relief effort. This application is very powerful because a user can communicate with all the other users of the network. To detect the misuse and abuse of this system, a strong implementation of non-repudiation is required. All the messages sent during the ad hoc phase can be analyzed once the network returns to the centralized mode i.e. base stations start functioning. The main drawback of this approach is that the malicious nodes are detected only after the attack. This is clearly not enough defence against attacks which are meant to cause massive destruction and loss of lives. To prevent these attacks from occuring, a mechanism is needed to revoke the malicious nodes as soon as they are detected. We solve this problem using a reputation based scheme where the malicious nodes are detected and revoked. This scheme is built on top of the security establishment scheme presented in Chapter 4.

The protocol assumes that each node has a private key corresponding to its ID in the ad hoc mode. We want to emphasize here that the ID of a node in the ad hoc mode is not same as its ID in the cellular mode. This is done to ensure anonymity in the ad hoc mode. This also prevents adversaries from mounting attacks aginst specific nodes of the network. If a node broadcasts messages using a known identity, the nodes location privacy may be compromised. These problems are resolved when we use separate ID's for each node. These ID's are assigned by the base station

and only the base station is able to derive the real ID of a node from the ad hoc ID.

The protocol starts with the base station assigning each node an ID, private key corresponding to the ID. Messages signed using the public key can be verified using the ID. The ID's are divided into groups such that each group has an ID which is the function of the ID's of all the nodes in the group. The private key corresponding to the group ID is split into n shares of which any k shares could reconstruct the group key successfully. For each node all the members in its group are its anonymous certifiers. Each node is given its reputation score in the form of a certificate which has the its ID and score. The certificate is signed by the private key of its group. Anyone with the ID of the node (from which the ID of the group can be derived) can verify the score of a node.

After the nodes enter the ad hoc mode, the broadcast messages to the entire network. Whenever the receiver of a broadcast message believes that the sender of the message is malicious, the receiver sends a negative recommendation to all the autonomous certifiers. When the current certificate of the sender expires, it would need a new certificate from the anonymous certifieres. The new certificate would reflect the negative recommendations for the node and reduce the reputation score of the node. If the reputation score of a node goes below a certain threshold, the anonymous certifiers revoke the node by broadcasting the revocation message in the network. The size of each group is made large enough that each node has atleast k nodes in its communication range of about 10 sq miles[33]. Once a certificate is generated by any k nodes of the autonomous group, it is transmitted to the other nodes of the autonomous group.

The nodes save the messages that they have sent and received in the ad hoc mode. After the base station is revived, all nodes send their messages to the base station

for audit. The base station performs and audit of the messages to ascertain if any of the nodes misbehaved. This audit would determine if any nodes attempted to spread misinformation against the honest nodes or sent positive recommendations for the malicious nodes.

For clear illustration we divide our scheme into 3 stages which are pre ad hoc mode, ad hoc mode and post ad hoc mode. We describe each of these stages in detail.

- **Pre Ad Hoc Mode:** In this stage the nodes are connected to the base station. This stage is performed at regular intervals to provide the latest information for the nodes in the ad hoc mode. The frequency with which the information in the nodes is updated is a tradeoff between overhead and accuracy. In this stage each node receives a reputation score and a share of the private key of the group to which it belongs along with its ID and private key [21]. The ID of each node is a combination of the group ID and the node ID. We represent the ID of a node P with ID_P and the group ID is represented by G_P. The private key corresponding to any ID sent by the base station is represented by SK_{ID}. SK_{G_P}/n is 1 share out of n for the reconstruction of the private key of the group. The communication from the base station can be represented as BS.

$$BS \rightarrow P : K_s \left[ID_P || SK_P || SK_{G_P}/n | \, | SCORE_P \right] \tag{5.1}$$

where

$$SCORE_P = SK_{G_P}[P||Value||timestamp] \tag{5.2}$$

The frequency at which this information is updated is a design parameter for the network administrators.

- **Ad Hoc Mode:** The nodes enter this stage when they fail to connect to the base station. Nodes can broadcast messages in this medium using the private keys that they received in the previous stage[21]. When the node P wants to broadcast in this environment it signs the message with its public key and sends the score certificate along with the broadcast.

$$Sender : SK_{ID}\left[M, ID\right] ||ID||SCORE_{ID} \tag{5.3}$$

The receivers are able to verify the signature on the message using the ID of the node and they are able to verify the score using the ID of the group to which the sender belongs.

$$Receiver : PK_{ID}\left[SK_{ID}\left[M, ID\right]\right] \tag{5.4}$$

$$Receiver : PK_{G_{ID}}\left[ID||Value||timestamp\right] \tag{5.5}$$

Once a receiver believes that a sender is malicious, the receiver sends a negative recommendation to all the anonymous certifiers of the sender. These nodes are all the nodes in the senders group excluding the sender. When the current certificate of the sender expires, the anonymous certifiers compute a new reputation score based on all the negative recommendations and collaboratively sign the new score with the private key of the group to which the sender belongs. Let R be the receiver and S be the sender.

$$R \rightarrow G_S : SK_R\left[RECO_R\right] ||ID_R \tag{5.6}$$

The Recommendation itself contains the ID of the recommender, ID of the recommended and the type of recommendation. It is then signed by the private key of the recommender. This is done to ensure the authenticity of the recommender and the recommended. The type of recommendation can

have the different levels of positive and negative recommendations that a recommender can make.

$$RECO_R = SK_R\,[ID_R||ID_S||Value] \tag{5.7}$$

We understand that the distributed nature of these networks may mean that the recommendations at all the anonymous certifiers may not be the same. Whenever an anonymous certifier reaches a threshold for the negative recommendations, it starts the process of getting the particular node revoked. If there are enough anonymous certifiers which believe that a node needs to be revoked, they can generate a revocation message. There may be a situation where the malicious node broadcasts only one malicious message after which it may not broadcast any other message. To prevent it from broadcasting other messages, the anonymous certifiers can issue no certificate in the future, but that will not discredit the message sent with a malicious certificate. This can be prevented by the anonymous certifiers by broadcasting the revocation certificate incase the reputation score goes below a threshold.

There are many schemes in literature which implement the distributed generation of a signature without any one of the collaborators knowing the signature[16, 17]. One of these schemes can be used by the anonymous certifiers to generate the private key of the group without any of the certifiers knowing the actual key.

- **Post Ad hoc Mode:**Nodes enter this stage when they are able to connect to the base station after the ad hoc mode. In this stage all the messages sent and received in the ad hoc mode by each node are audited. This may also involve the use of the law enforcement agencies to investigate users of the cell phones for the messages that were sent out during the ad hoc mode.

This stage may be used for correcting any of the trust scores set during the ad hoc mode. Nodes may be rewarded or punished based on their behavior in the ad hoc mode. The recommendations made by the nodes may also be analyzed to determine if a given node is a collaborator of the malicious node. Such nodes may be revoked in the centralized system which is trivial. The post ad hoc node also gives the network administrators an opportunity to correct any mistakes like the revocation of honest nodes, which may have been made in the ad hoc mode. This stage helps in obtaining better reputation scores for nodes when the enter the ad hoc mode in the future.

5.5 Evaluation

In this section we evaluate the protocol through mathematical analysis and simulations. WE first list all the terminology and parameters used in this section. The values of these parameters would influence the security and overhead of this scheme.

- **Group Compromise:** All the nodes of the network are divided into groups. For each node of the group, all the other nodes act as the anonymous certifiers. If the number of malicious nodes in a group is more than the number of nodes required to reconstruct the private key of the group, then the group is said to be compromised.

- **Node Compromise:** When a group is compromised, all the nodes in the group which are not malicious can be revoked. We refer to this as node compromise.

- **Area of Network Deployment:** This is the total area across which the network is deployed. We assume that the communication range of a node is

equal in all directions. We represent the deployment region as a circle for clear illustration.

- **Group Size:** It is the number of elements in each group. This is a critical design parameter because for the same level of threshold, if the size of the group is small the probability of there existing a malicious node in the group is also small. On the other hand a small group size reduces the possibility of each node having enough certifiers in its neighborhood.

- **Threshold:** It is the number of anonymous certifiers required to reconstruct the certificate of the group. A lower value of threshold makes the group vulnerable to compromise by a small number of malicious nodes whereas a large value of threshold increases the overhead required for generating a signature. It is also much harder for a node to find enough certifiers to refresh its score certificate.

- **Denial of Service:** Threshold cryptography works when the number of malicious nodes is less than the threshold. But, if the number of malicious nodes in threshold cryptography is large enough such that they are able to prevent the signing of score certificates then the whole system would become dysfunctional. For a (n, k) threshold scheme, the number of nodes required to sign the certificate is k but the number of nodes required to prevent the signing of the certificate is $(n - k)$.

- **Time for Certificate Renewal:** The time for which a certificate is valid is an important design issue. The lifetime of a certificate presents a tradeoff between security and overhead. If the lifetime of a certificate is small, the overhead for signing the certificates on the network would be larger whereas

security would be better because the number of broadcasts which can be made by a node is limited.

5.5.1 Theoretical Analysis

We now analyze the probability of a group being compromised which is equivalent to the probability of k out of n nodes of the group being compromised. Let the total number of nodes in the network be N with α being the total malicious nodes in the network. The probability of a particular node being malicious is $\frac{\alpha}{N}$ and the probability of a node belonging to a particular group is $\frac{n}{N}$. The probability of a node being malicious and belonging to a particular group is the product of the two probabilities which is $\frac{n\alpha}{N^2}$. Therefore the probability of k nodes of a group being compromised is

$$P_C = \left(\frac{n\alpha}{N^2}\right)^k \tag{5.8}$$

This equation presents some interesting results. If the number of nodes in the network increases, the probability of a malicious node belonging to a particular group decreases which reduces the probability of the group being compromised. An increase in the value of group size would mean that the malicious groups are distributed across fewer groups as a result of which the probability of a group being compromised increases. The most important factor that influences group security is the threshold. A higher threshold would reduce the probability of group compromise exponentially.

We now analyze the problem of Denial of Service in a group. If the number of malicious nodes in a group is less than the threshold value, they will be unable to sign the certificates. But, they can cause DoS by not participating in the signing of score certificates of other nodes of the network. For an (n, k) threshold scheme, if the number of malicious nodes α is greater than $(n - k)$, the group is vulnerable

to DoS because without the participation of malicious nodes the honest nodes of the network can not create signatures. To make this scheme robust to DoS attacks the following conditions have to be satisfied.

$$\alpha \leq (n - k) \tag{5.9}$$

The best possible robustness against DoS is achieved when the group is resilient against $k - 1$ malicious nodes. If $\alpha = k - 1$ then

$$k - 1 \leq n - k \tag{5.10}$$

$$\Rightarrow k \leq \frac{n - 1}{2} \tag{5.11}$$

If the value of α is greater than q, then the group is compromised.

Let the deployment region have a radius of R and each node have a communication range of r. We assume that the nodes of a group are deployed uniformly throughout the network. Let the density of nodes be ρ. For an (n, k) threshold scheme, the minimum number of nodes within the communication range of a node is k. Therefore the minimum density of nodes required is

$$\rho = \frac{k}{\pi r^2} \tag{5.12}$$

If the total area of the network is πR^2, the minimum number of nodes required for each group n would be

$$n = k \left(\frac{R}{r} \right)^2 \tag{5.13}$$

With the increase in the area of deployment, the number of nodes required in each group increases. We believe that the area of the deployment region would be comparable to the communication range of the nodes.

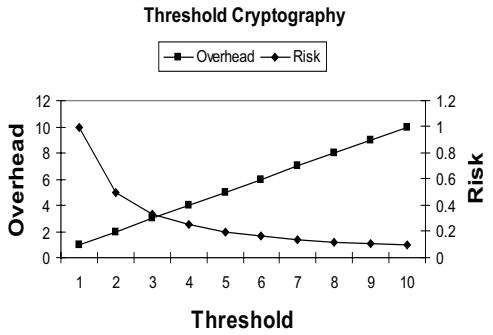


FIGURE 5.1. This figure shows the relation betweeen threshold, overhead and risk in threshold cryptography. As the threshold increases the risk decreases and overhead needed to reconstruct the secret increases

5.5.2 Simulations

We have also simulated the different aspects of this paper. The simulations bring out the relations between the different aspects of this paper.

In Figure 5.1 we present the basic principle of threshold cryptography. It shows the relation between threshold, overhead and risk. As the threshold increases, the overhead required to reconstruct the shared secret increases. Each node has to generate a partial signature to the score certificate which is then combined to generate the final score certificate. As a result overhead is directly proportional to the threshold. Risk on the other hand is inversely proportional to threshold. If the threshold is 1, the risk is the highest. As the value of threshold increases, the number of nodes which are required to sign the certificate increases. This is clearly illustrated in Figure 5.1. In this figure we assume that each node of the network has only share of the group private key. If the use of some pre trusted nodes like cops are made, the risk of threshold cryptography can be lowered.

In Figure 5.2 we show the relationship between the network size and the number of nodes which are within communication range. The communication range of each node is assumed to be constant. The area of deployment is assumed to be circular

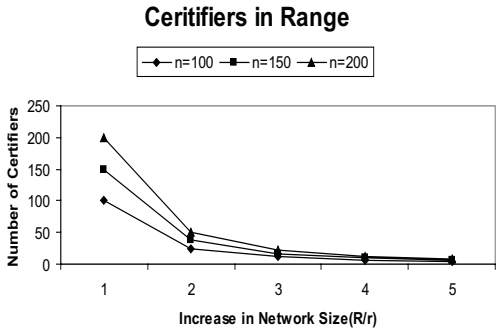


FIGURE 5.2. This figure shows the relation betweeen the number of anonymous certifiers within communication range and the area of network deployment. r/R indicates the ratio of the radius of the deployment region to the communication range of a node

and represented in terms of the communication range of each node. This simulation assumes that there are 10000 nodes in the network and the no. of anonymous certifiers are distributed uniformly throughout the network. We can clearly see that with the increase in the area of the network, the number of anonymous certifiers decreases dramatically. We believe that in a real world implementation the sizes of the groups would be very large compared to the threshold size to account for a reasonably large deployment area. We also believe that the size of the ad hoc network would be comparable to the communication range of the nodes.

In Figure 5.3, we present the relationship between the number of nodes compromised when a certain number of nodes are malicious for different values of group size. A node is compromised when atleast k nodes of its group are malicious. In this simulation, the number total number of nodes is taken as 10000. We perform this simulation for different values of the group sizes. All the values in Figure 5.3 are represented as a percentage of the total nodes in the network. When the number of malicious nodes is small, an increase the number of malicious nodes does not compromise honest nodes because the number of malicious nodes in all groups is less than the threshold. As the number of malicious nodes increases, the number

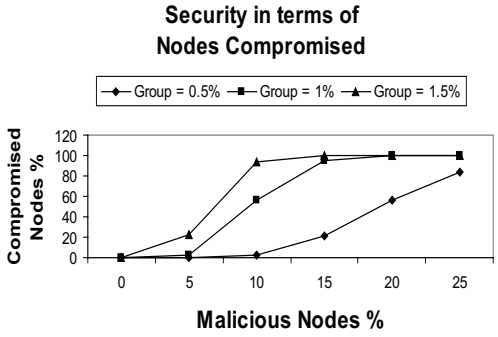


FIGURE 5.3. An increase in the no. of malicious nodes increases the number of nodes compromised.An increase in group size causes the number of malicious nodes are spread across fewer groups. Hence the rapid increase in nodes compromised. Values are represented as fractions of total nodes.

of groups in which the number of malicious nodes is greater than the threshold increases suddenly. This figure shows that as long as the number of malicious nodes is small compared to the total nodes (around 1-2%), the number of nodes compromised is negligible. Figure 5.3 also shows that an increase in group size increases the number of groups compromised for a given level of malicious nodes. This behavior is expected because when the group size is increased and the threshold is kept constant, the malicious nodes are spread across fewer groups. As a result the number of malicious nodes in each group is higher and the number of malicious nodes reaches the group threshold faster.

In Figure 5.4, we present the relationship between the nodes compromised and the threshold. Intutively, an increase in the value of threshold would result in better security. Hence, fewer compromised nodes. In this figure, threshold is represented as a fraction of the group size whereas the nodes compromised and the malicious nodes are represented as fractions of total nodes in the network. For this simulation, the group size was taken as 100 and the network size was taken as 10000. For a high value of malicious nodes at 5%, a threshold of 15% of group size achieves zero

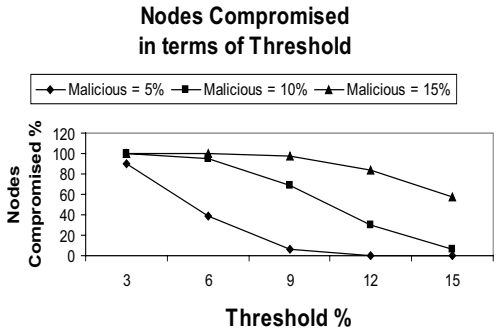


FIGURE 5.4. This figure shows the relation between the nodes compromised and the threshold. As the threshold increases the value of nodes compromised decreases. The nodes compromised and malicious nodes are a fraction of the total nodes of the network whereas the threshold is a fraction of group size.

compromised nodes. We believe that in such networks the malicious nodes would be a negligible when compared to the total nodes in the network.

Chapter 6

Secure Spatial Authentication using Cell Phones

6.1 Introduction

The capabilities of cell phones have increased dramatically over the last few years. In addition to the standard telephone features, the phones also Instant Messaging, MMS, Internet access etc. More advanced features like music and video streaming, digital camera, document scanner are being bundled with the cell phone. These features have transformed the cell phone from a simple phone to a digital swiss army knife.

More advanced features like bluetooth, IR have been added to allow the cell phone to connect with other devices. Avaya, Motorola, and Proxim are planning to introduce a new class of mobile phones called the enterprise phones [2, 3], also referred to as the dual phones. These phones will be able to make voice calls over the cellular network and the 802.11a WLAN networks. The advantage of using this phone is that the user can make calls through the WLAN infrastructure when he is able to connect to the WLAN. This would save money because the cell phone user would be able to use the WLAN minutes for free. The companies have also developed the technology to hand off calls between the WLAN and cellular network. The ability of these phones to connect to both the local WLAN and the cellular network allows the use of these dual phones for many different applications.

There are many applications in wireless networks where access is granted to a user only when the user is located in certain predefined locations [26, 40, 55]. For eg. a doctor should be able to access the medical records only when he is located inside the hospital and not in cafeteria. In this scenario the doctor has access to

the medical records only when he is located in a safe place like his office and not in a public place like the cafeteria. The server can be sure of the users location by using a trusted hardware sensor which is able to determine if the cell phone is in its communication range. Another approach to be certain of the location of the phone is to have a GPS module on the SIM card. To lie about his location, the user of the cell phone would have to make a fake SIM card which is extremely hard. This is the most basic assumption on which the entire cell phone industry is based.

Dual phones can be used for authentication in many other situations [23]. Access to buildings, offices, labs are controlled by RFID enabled access cards. These access cards work when they are placed close to the RFID reader. The problem with such a scenario is that the user has to have a separate access card for each location. Moreover, the signal coming out of the access card is the same all the time, which makes it vulnerable to duplication. We believe that dual cell phones coupled with the strong cellular network would provide better security than the use of use of cheap RFID based access cards. When one of these phones would come into contact with the access servers and request for authentication, the server sends a random challenge to the phone through the cellular network. This challenge is retransmitted to a secure hardware sensor connected to the hardware server. This approach verifies both the identity and the location of the user. Another approach to verify the location of the user is through the use of GPS. When the cell phone wants to get authenticated it sends the random challenge obtained from the authentication server through the cellular network along with the GPS coordinates. Our scheme is clearly illustrated through Figures 6.1 and 6.2.

We list some of the advantages of using dual cell phones over the traditional access cards.

- **Ease of Use:** A single cell phone can replace all the access cards required by a user. To add a new user, the authentication server has to store the phone number corresponding to every user. Incase the cell phone is lost, the base station can revoke the phone. This revokes the user from all the authentication servers.

- **Security:** The security offered by these devices is much stronger than the traditional access cards. Unlike the access card, it is much harder to fake the SIM card of a cell phone. Moreover the authentication server provides a new random challenge everytime the client requests for access. As a result the challenge is impossible to copy and reuse.

- **Spatial Control:** Since the random challenge sent to the phone by the authentication server through the cellular network has to be presented to the authentication server using WLAN , it makes it very hard for the user to present false location information.

We believe that it would be very hard to duplicate or manipulate the SIM card of a cell phone. As all the information and code required for authentication would be coded onto the SIM, the difficulty in compromising the SIM card would help in ensuring spatial control. Also, this system would work without any human intervention.

A potential problem for a cellular network based authentication system is that its working is dependent on the proper functioning of the cellular network. The network could assign a higher priority to such authentication packets and provide more stringent QoS requirements. There are other tradeoffs between security and overhead. We explore these tradeoffs in section 6.3.

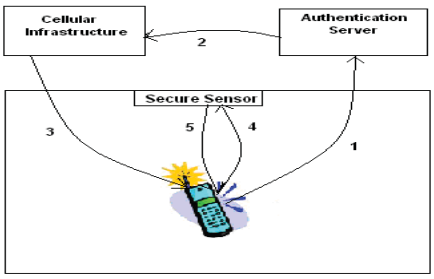


FIGURE 6.1. This figure shows the different stages in this scheme. In stage 1 authentication is requested. In stage 2 the authentication server verifies the authenticitity of the user and sends a random challenge to the user through the cellular infrastructure. In stage 3, the cell infrastructure sends this random challenge to the cell phone. In stages 4 and 5 the cell phone proves to the authentication server that it is indeed seeking authentication and located in a appropriate location.

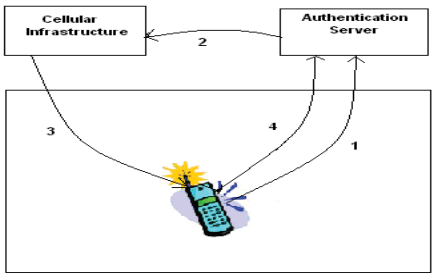


FIGURE 6.2. This figure shows the different stages in this scheme when GPS is used. In stage 1 authentication is requested. In stage 2 the authentication server verifies the authenticitity of the user and sends a random challenge to the user through the cellular infrastructure. In stage 3, the cell infrastructure sends this random challenge to the cell phone. In stages 4, the cell phone sends the random challenge and the GPS coordinates to the Authentication

6.2 Related Work

Two factor authentication requires that a client produce two independent factors based on which a server can verify its authenticity [39, 54]. These two parameters usually are "something that you know" and "something that you have". It has been extensively implemented using Smart Cards and USB Tokens[54]. Popular two- factor authentication schemes include the SecurID token by RSA Security and similar products by VeriSign,ActivPack,SafeNet,CRYPTOCard,Rainbow Technologies and others. These products have a backend server and a token carried by the user. This token generates a random one time password using some pseudo random paramteres. This is combined with the password known to the client to create the password dependent on two factors.

Two factor authentication has also been implemented using the cell phones as the second factor. When the user requests for an authentication, the authentication server sends a random passcode to the users cell phone [58]. This combined with the password known to the user makes the one time password required in two factor authentication. This scheme is offered by products like RSA Mobile.

ASB Bank in New Zealand has implemented a scheme called Netcode Authentication System. It uses the cell phone as the token in a two factor authentication system. When the user wants to perform a transaction, the system sends the user a 8 digit authetication code to the cell phone through a text message. The user of the phone then has to enter this number back into the computer to verify the authenticity of the user.

The present scheme can only be used for authentication but can not be used for spatial-control. We present a scheme in the following section which extends the use of dual cell phones to spatial-control.

6.3 Our Scheme

Our scheme has two parts namely user authentication and spatial control. In the authentication part, the authentication server verifies if the cell phone is genuine or not by sending a message to the cell phone. If the owner of the cell phone is able to reproduce the message sent by the authentication server to the cell phone, then the phone is considered to be genuine. Otherwise the phone is not authenticated.

6.3.1 User Authentication

Once the phone is deemed to be genuine, the authentication server would like to enforce the spatial control. This is useful when the services being provided to the base station are dependent on the location of the cell phone. For e.g. even though a doctor has access to some health records, he should be allowed to access those records in his office and not in the cafeteria. This problem is non-trivial if the doctor is allowed to access the records through multiple computers at multiple locations. We address this problem by verifying the location of a cell phone using another trusted hardware sensor which is deployed in all the locations at which access is to be granted to the user. We also propose to use GPS to determine the location of a phone.

We now present the different stages of the authentication process.

- **Authentication Request:** In this stage the WLAN portion of the dual phone transmits its ID to the authentication server. The phone shares a common ID for both the cellular network and the WLAN.

- **Authentication Reply:**Once the authentication server receives a request, it checks if the user corresponding to the ID is authorized. If the user is valid, it generates a random challenge and sends it to the user through the cellular

network. This challenge would reach the cell phone only if it has not been revoked at the base station.

6.3.2 Spatial Authentication

Once the authentication is performed, the authentication server has to check if the cell phone is in the desired location. To detect the presence of the cell phones, we place a hardware sensor in every location where access to the cell phone is allowed. Once the phone receives the random challenge from the authentication server via the base station, the WLAN portion of the cell phone sends this random challenge to the hardware sensor which is securely connected to the authentication server. The hardware sensor then verifies the random challenge. If the random challenge sent to the cell phone through the base station matches the challenge sent to the trusted sensor, the user is authenticated and the server is able to spatial control because the cell phone is close to one of the trusted hardware sensors. This approach is clearly illustrated in Figure 6.1

Another possible approach for a server to establish spatial control is to use GPS[14]. When a cell phone seeks authentication, it sends the GPS coordinates along with the random challenge received from the authentication server. We believe that the GPS module in the phone would be coded into the SIM card. As a result it would be very hard to give false GPS values because that would require building a fake SIM card which has everything same as the original SIM except the GPS receiver. This approach relies on the fundamental assumption that it is hard to duplicate a SIM card.

The major advantages of using the trusted hardware sensor is the accuracy and the reliability that offers. This system is more reliable and consistent because it does not depend on the vagaries of the GPS system. The disadvantage of this approach is the extra cost and effort required to deploy and maintain the trusted

hardware sensors. Moreover, this approch fixes the number of locations where the cell phone user can get authenticated and prove his location to the authentication server. This approach is unlike the GPS based approach which aviods the overhead of deploying and managing the trusted hardware sensors. The disadvantage of using the GPS based approach is the inconsistency in the behavior of GPS. To address the inconsistency in the behavior of GPS many approaches like Differential GPS have been proposed [24].

6.3.3 Mobility Management

Once a cell phone gets authenticated and proves its location to the authentication server, it should not be allowed to change its location and move to an inappropriate location. To remove this vulnerability, the proposed scheme should verify the location of the cell phone at regular intervals. In the trusted sensor based scheme, the sensor could verify the existance of the cell phone at the location at regular intervals. In a GPS based scheme, the authentication server could request for the GPS coordinates of the cell phone at regular intervals. These methods allow the authentication server to continuously establish spatial control over the cell phone.

In our scheme with trusted sensors, once the phone gets authenticated and gets a random challenge from the authentication server, the WLAN portion of the dual phone sends this information to the trusted sensor. This not only authenticates the cell phone to the authentication server but also proves its location. There could be many locations where the cell phone may be authorized to receive service. A malicious user could gain access through the authentication server by replicating the signal sent by the cell phone to the trusted sensor. We present some solutions which can be used to tackle this problem. These solutions offer different degrees of overhead and security.

- **Multiple Logins:**If the authentication server receives the same random challenge from multiple locations, the same cell phone is being used login to the system through multiple locations. The authentication server could prevent this attack by storing all the random challenges and comparing them to the challenges presented by different cell phones. This approach would fail when a user gets authenticated by the system and leaves without loging out because it would allow the malicious node to login to the system by faking the random challenge.

  To resolve these issues we propose to use multiple random challenges for one user authentication session. When the hardware sensor requests for a challenge, the cell phone provides one of the several random challenges that it received from the authentication server. This prevents the reuse of the same random challenge and once the cell phone exhausts all the random challenges, it has to reauthenticate itself using the cellular infrastructure.

- **Mobility Control:**Mobility control is an important aspect of this scheme. One approach is to allow a authenticated cell phone mobility across all appropriate locations. A diametrically opposite approach would be to authenticate the cell phone at every location. The latter approach would provide much better security with the overhead of using the cellular infrastructure to authenticate the cell phone. A tradeoff between the two approaches would be to allow access to locations geographically close to the location at which the authentication has been done. This would be a design parameter for the network administrator.

  All the 16 locations in Figure 6.3 are valid for a given cell phone. In the first approach where the cell phone has to get authenticated once for all

1	**2**	**3**	**4**
5	**6**	**7**	**8**
9	**10**	**11**	**12**
13	**14**	**15**	**16**

FIGURE 6.3. This figure shows a sample deployment region consisting of 16 valid locations.

locations, the cell phone can get authenticated in any location from 1 to 16 and use the serveices at all locations. In the second case, the user requires authentication at all locations. For e.g. the user would require authentication through the cellular network to move from cell 2 to cell 3. In the final case, once the user is authenticated in a particular location, he does not require to get authenticated in the cells which are close to his current cell. For e.g. if a user is authenticated in cell 1, then he does not have to obtain authentication to get the service from cells 2, 3, 4.

6.4 Analysis

In this section we analyze the performance and the tradeoffs involved in the authentication and spatial control. Once the authentication server determines that a given cell phone is genuine and it is located in the appropriate place, access is granted. Once the access is granted the client should not be in a position to move into an inappropriate location and continue to retain access to the services. To prevent this problem, the location of the device needs to be constantly monitored. The frequency at which the device is monitired presents an interesting tradeoff. If

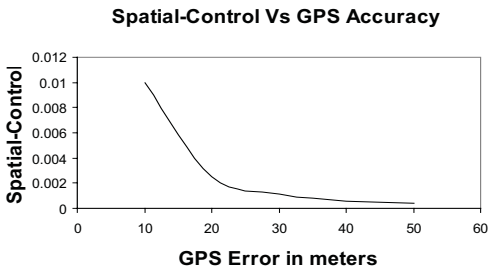


FIGURE 6.4. This figure shows the tradeoff between frequency of authentication and overhead

the frequency of monitoring is high, the guarantee of the cell phone being in the appropriate location is high. On the other hand if the frequency of this monitoring is low, the overhead is low and the probability of the cell phone being in a inappropriate location is high. For this simulation we consider interval of verification in minutes and the overhead is measured in terms of messages per minute using the cellular infrastructure. This is captured in Figure 6.4.

Although the overhead for the scheme using GPS receivers is low compared to the scheme using a trusted sensor, the varying accuracy of GPS may lead to difficulties in spatial control. The coordinates shown by the receiver may vary from the actual coordinates. The decrease in the accuracy for GPS would result in lesser spatial control. In fig.2 we define spatial control as the probability of obtaining the exact coordinates of a location from GPS. For e.g. if the GPS error is 10 units, the probability of the GPS finding the exact location is given by $\frac{\pi(1)^2}{\pi(10)^2}$. For this simulation the GPS accuracy is measured in meters and the spatial control is the prabability of the GPS values being accurate. The relation between GPS error and spatial control is clearly shown in Figure 6.5.

Figure 6.6 we show the overhead on the cellular network for different levels of security for mobility management. We consider three cases for the overhead analysis. The first case is Single location access where a cell phone requires authentication

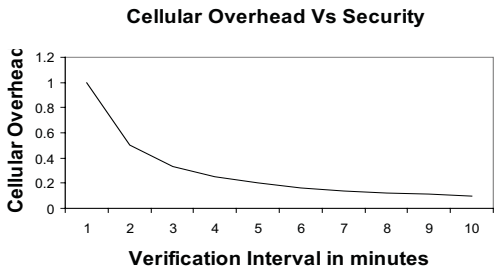


FIGURE 6.5. This figure shows the relation between GPS error and spatial control

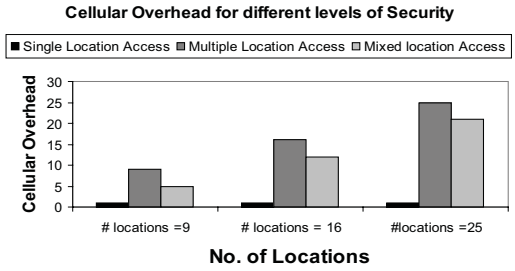


FIGURE 6.6. This figure shows the relationship between overhead and security. For single location access one authentication is required to access all the locations which results in least overhead where for multiple location access separate authentication is required for each cell resulting in extremely high overhead. A tradeoff between these two approaches is shown in the mixed location access.

only once. The second case is the Multiple Location Access where the cell phone requires authentication at all locations. The third case is the Mixed location access where a cell phone does not require to obtain authentication in 4 of its neighboring cells of Figure 6.3.

Chapter 7
Conclusions

There is a tradeoff among security, connectivity and cost in sensor networks. Through analysis and simulations we showed that our proposed management scheme SCON enables good tradeoff among security and connectivity by making use of nodes of previous deployments. SCON makes use of strong node and actors when they are available, but does not count on them. The architecture is incremental because SCON works with or without these special nodes. SCON can be used as a management solution for multiple deployments with most of the existing key predistribution schemes. Sensor networks with heterogeneous nodes have a wide range of applications. These applications need to establish secure connectivity between the mobile and the stationary nodes of the network. The mobile nodes may need unrestricted movement through different sensor networks. The existing key predistribution schemes restrict the mobility of the nodes to only one network. We present two schemes namely, key predistribution using separate key pool and key predistribution using segmented key pool. They allow the mobile nodes to interact with the stationary nodes of different networks.

We propose to use cell phones as nodes of an ad hoc network when they are unable to connect to the base station. These omnipresent devices can have many applications for emergency situations and disaster management. Both resource constrains and the nature of traffic make this emergency application of cell phones very unique and, consequently, new communication and security protocols are needed. We also present a scheme to ensure non-repudiation of messages in this environment. Using a fixed subset of nodes as anonymous certifiers, we also provide a

distributed reputation based scheme for node revocation in ad hoc networks. Many other interesting applications of cell phones in the dual mode have been provided which use to robust cellular network for authentication and spatial authentication.

References

[1] Cellularonline, http://www.cellular.co.za.

[2] http://www.networkworld.com/news/2004/072604avaya.html.

[3] http://www.pcworld.com/news/article/0,aid,116334,00.asp.

[4] G. N. Aggelou and R. Tafazolli. On the Relaying Capacity of Next-generation GSM Cellular Networks. *IEEE Personal Communications Magazine*, 8(1):40–47, 2001.

[5] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci. Wireless sensor networks: A survey. *Computer Networks*, 38(4):393–422, 2002.

[6] Ian F. Akyldiz and Ismail H. Kasimoglu. Wireless sensor and actor networks: research challenges. *ADHOC Networks*, 2004.

[7] Rafael Ballagas, Michael Rohs, Jennifer Sheridan, and Jan Borchers. The smart phone: A ubiquitous input device. *IEEE Pervasive Computing*, 5(1):70–77, Jan-Mar 2006.

[8] Dan Boneh, Ben Lynn, and Hovav Shacham. Short signatures from the weil pairing. *J. Cryptol.*, 17(4):297–319, 2004.

[9] Levente Buttyan and Jean-Pierre Hubaux. Stimulating cooperation in self-organizing mobile ad hoc networks. *Mob. Netw. Appl.*, 8(5):579–592, 2003.

[10] Jian Cai and David J. Goodman. General packet radio service in gsm. *IEEE Communications Magazine*, pages 122–131, October 1997.

[11] D. W. Carman, P. S. Kruus, and B. J. Matt. Constraints and approaches for distributed sensor network security. Technical Report 00-010, NAI Labs Technical Report, September 1,2000.

[12] H. Chan, A. Perrig, and D. Song. Key distribution techniques for sensor networks. *Wireless Sensor Networks ISBN:1-4020-7883-8*, pages 277–303.

[13] Haowen Chan, Adrian Perrig, and Dawn Song. Random key predistribution schemes for sensor networks. *IEEE Symposium on Security and Privacy*, pages 197–213, May 11–14,2003.

[14] Peter H. Dana. Global positioning system (gps) time dissemination for real-time applications. *Real-Time Syst.*, 12(1):9–40, 1997.

[15] Carlton R. Davis. A localized trust management scheme for ad hoc networks. *Proceedings of 3rd International Conference on Networking (ICN'04)*, pages 671–675, Mar 2004.

[16] Yvo Desmedt and Yair Frankel. Shared generation of authenticators and signatures (extended abstract). In *CRYPTO '91: Proceedings of the 11th Annual International Cryptology Conference on Advances in Cryptology*, pages 457–469, London, UK, 1992. Springer-Verlag.

[17] Yvo G. Desmedt and Yair Frankel. Threshold cryptosystems. In *CRYPTO '89: Proceedings on Advances in cryptology*, pages 307–315, New York, NY, USA, 1989. Springer-Verlag New York, Inc.

[18] W. Diffie and M. E. Hellman. New directions in cryptography. *IEEE Transactions on Information Theory*, 22:644–654, 1976.

[19] Wenliang Du, Jing Deng, Yunghsiang S. Han, Shigang Chen, and Pramod K. Varshney. A key management scheme for wireless sensor networks using deployment knowledge. *In Proceedings of the IEEE INFOCOM'04*, pages 586–597, March 7–11,2004.

[20] A. Durresi, V. Paruchuri, R.Kannan, and S.S.Iyengar. Optimized broadcast protocol for sensor networks. *IEEE Transactions on Computers*, 54(8):1013–1024, August 2005.

[21] Arjan Durresi, Vijay Bulusu, and Vamsi Paruchuri. Security in ad hoc networks on cellular phones for emergency situations. *Accepted in Ad Hoc Networks Journal*, 2006.

[22] L. Eschenauer, V. Gligor, and J. Baras. On trust establishment in mobile ad-hoc networks, 2002.

[23] Alberto Escudero-Pascual. Privacy enhanced architecture for location based services in the next generation wireless networks. 2002.

[24] Jay Farrell and Tony Givargis. Differential gps reference station algorithm: Design and analysis, 2000.

[25] Alberto Fernandes, Evangelos Kotsovinos, Sven Ostring, and Boris Dragovic. Pinocchio: Incentives for honest participation in distributed trust management. In *Proceedings of the 2nd International Conference on Trust Management (iTrust 2004)*, pages 63–77, Oxford, UK, March 2004. Also published in Springer-Verlag Lecture Notes in Computer Science (LNCS), Volume 2995, pp. 63-77.

[26] Frode Hansen and Vladimir Oleshchuk. Application of role-based access control in wireless healthcare information systems. In *Proc. For Scandinavian Conference in Health Informatics*, pages 30–33, 2003.

[27] J. Horwitz and B. Lynn. Toward hierarchical identity-based encryption. *Proc. of Eurocrypt02, LNCS 2332,Springer-Verlag*, pages 466–481, 2002.

[28] D. B. Johnson and D. A. Maltz. Dynamic Source Routing in Ad Hoc Wireless Networks. *Mobile Computing. Kluwer Academic Publishers*, 353:153–181, 1996.

[29] Rajgopal Kannan, Lydia Ray, Arjan Durresi, and S.S.Iyengar. Security-performance tradeoffs of inheritance based key presistribution for wireless sensor networks. *CoRR ArXiv:Computing Research Repository*, June 2004.

[30] Chris Karlof and David Wagner. Secure routing in wireless sensor networks: attacks and countermeasures. *First IEEE International Workshop on Sensor Network Protocols and Applications*, May 2003.

[31] M. Kinateder and K. Rothermel. Architecture and Algorithms for a Distributed Reputation System. In P. Nixon and S. Terzis, editors, *Proceedings of the First International Conference on Trust Management*, volume 2692 of *LNCS*, pages 1–16, Crete, Greece, May 2003. Springer-Verlag.

[32] John Kohl and B. Clifford Neuman. The kerberos network authentication service (v5). *RFC 1510*, September 1993.

[33] Julia Layton, Marshall Brain, and Jeff Tyson. Introduction to how cell phones work, 2005.

[34] Donggang Liu and Peng Ning. Location-based pairwise key establishments for static sensor networks. *ACM workshop on Security in Ad Hoc and Sensor Networks*, 2003.

[35] Donggang Liu, Peng Ning, and Rongfang Li. Establishing pairwise keys in distributed sensor networks. *10th ACM conference on Computers and Communication Security (CCS 03)*, pages 52–61, October 2003.

[36] Jinshan Liu and Valérie Issarny. Enhanced reputation mechanism for mobile ad hoc networks. In *iTrust*, pages 48–62, 2004.

[37] C.C. Lo and Y.J. Chen. Secure communication mechanisms for gsm networks. *IEEE Trans. Consumer Electronics*, 45(4):1074–1080, November 1999.

[38] Haiyun Luo, Ramachandran Ramjee, Prasun Sinha, Li Li, and Songwu Lu. UCAN: A Unified Cellular and Ad-Hoc Network Architecture. In *Proceedings of MobiCom 2003*, pages 353–367, San Diego, CA, September 14-19, 2003.

[39] Philip MacKenzie and Mike Reiter. Networked cryptographic devices resilient to capture. Technical Report 2001-19, 2001.

[40] I. Mavridis, C. Georgiadis, and G. Pangalos. Access-rule certificates for secure distributed healthcare applications over the internet.

[41] A. Mehrotra and L.S. Golding. Mobility and security management in the gsm system and some proposed future improvements. *Proceedings of the IEEE*, 86(7):1480–1497, July 1998.

[42] S. P. Miller, C. Neuman, J. I. Schiller, and J. H. Saltzer. Kerberos authentication and authorization system. *In Project Athena Technical Plan*, page E.2.1, 1987.

[43] L. Mui, M. Mohtashemi, and A. Halberstadt. A computational model of trust and reputation for e-businesses. In *HICSS '02: Proceedings of the 35th Annual Hawaii International Conference on System Sciences (HICSS'02)-Volume 7*, page 188, Washington, DC, USA, 2002. IEEE Computer Society.

[44] P. Obreiter. A case for evidence-aware distributed reputation systems — overcoming the limitations of plausibility considerations. In *Second International Conference on Trust Management (iTrust'04), Oxford, UK*, 2004.

[45] C. E. Perkins and E. M. Royer. Ad-hoc On Demand Distance Vector Routing. In *Proceedings of IEEE Workshop on Mobile Computing Systems and Applications (WMCSA)*, pages 90–100, 1999.

[46] E.M. Petriu, N.D. Georganas, D.C. Petriu, D. Makrakis, and V.Z.Groza. Sensor based information appliances. *IEEE Instrumentation and Measurement Magazine*, 3(4):31–35, 2000.

[47] Report of NSF Wireless Mobile Planning Group (WMPG) Workshop. New Architectures and Disruptive Technologies for the Future Internet: The Wireless, Mobile and Sensor Network Perspective, August 2005.

[48] R. L. Rivest, A. Shamir, and L. M. Adleman. A method for obtaining digital signatures and public-key cryptosystems. *Communications of the ACM*, 21(2):120–126, 1978.

[49] E. Royer and C. K. Toh. A review of current routing protocols for ad hoc mobile wireless networks. *IEEE Personal Communication*, pages 46–65, April 1999.

[50] A. Seshadri, A. Perrig, L. van Doorn, and P. Khosla. Swatt: Software-based attestation for embedded device. *2004 IEEE Symposium on Security and Privacy*, March 9–12,2004.

[51] Adi Shamir. Identity-based cryptosystems and signature schemes. In *Proceedings of CRYPTO 84 on Advances in cryptology*, pages 47–53, New York, NY, USA, 1985. Springer-Verlag New York, Inc.

[52] J. Spencer. *The Strange Logic of Random Graphs ISBN: 3-540-41654-4*. Springer-Verlag, August 9,2001.

[53] Lakshminarayanan Subramanian, , and Randy H. Katz. An architecture for building self-configurable systems. *Proceedings of IEEE/ACM Workshop on Mobile Ad Hoc Networking and Computing*, August 2000.

[54] Rainbow Technologies. Two-factor authentication? making sense of all the options. Technical report, 2001.

[55] Eleanor Toye, Richard Sharp, Anil Madhavapeddy, and David Scott. Using smart phones to access site-specific services. *IEEE Pervasive Computing*, 04(2):60–66, 2005.

[56] Mohit Virendra and Shambhu Upadhyaya. Securing information through trust management in wireless networks. *Workshop on Secure Knowledge Management (SKM 2004), Buffalo, NY*, pages 201–206, Sep 2004.

[57] Anthony D. Wood and John A. Stankovic. Denial of service in sensor networks. *IEEE Computer*, 35(10):54–62, 2002.

[58] Min Wu, Simson Garfinkel, and Robert Miller. Secure web authentication with mobile phones. *MIT Computer Science and Artificial Intelligence Lab*, 2004.

[59] Yan Yu, Ramesh Govindan, and Deborah Estrin. Geographical and energy-aware routing: A recursive data dissemination protocol for wireless sensor networks. *UCLA Computer Science Department Technical Report UCLA/CSD-TR-01-0023*, May 2001.

[60] Lidong Zhou and Zygmunt J. Haas. Securing ad hoc networks. *IEEE Network*, 13(6):24–30, 1999.

[61] Sencun Zhu, Sanjeev Setia, and Sushil Jajodia. Leap: Efficient security mechanisms for large-scale distributed sensor networks. *10th ACM conference on Computers and Communication Security (CCS 03)*, pages 62–72, October 2003.

Vita

Vijay Bulusu received his Bachelor of Engineering in Information Technology from Indian Institute of Information Technology-Calcutta, India in May 2004. Subsequently, he worked as an associate consultant in Kanbay Corporation, Hyderabad, India for 6 months. He joined the Department of Computer Science, Louisiana State University, Baton Rouge in Fall 2004. His current research interests include security and QoS in wireless sensor networks, ad hoc networks, vehicular networks and cellular networks.